INDOOR AIR QUALITY, FUEL CHOICE, AND INFANT MORTALITY

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Tsenguunjav Byambasuren

August 2019

ABSTRACT

Indoor air pollution (IAP)–predominantly from solid fuel use for cooking–is a global health threat, particularly for women and young children, and one of the leading causes of infant deaths worldwide in developing and emerging countries. This paper estimates the causal effect of cooking fuel choice on infant mortality, focusing on children under 5 years of age, through the channel of IAP in India using National Family Health Survey (NFHS) data during the period 1992–2016. The main empirical framework quantifies how solid cooking fuel, an indirect measure of IAP, affects under-five mortality across all 36 states in the country during this period. To address the potential endogeneity issue in the relationship between type of cooking fuel and mortality, I instrument for cooking fuel choice using forest cover and agricultural land ownership, which induce significant variations in fuel type. The non-IV results show that the use of solid fuel for cooking increases the risk of mortality in children aged under five by 0.8 percentage points, indicating that previous studies over-estimated the marginal impact by about 0.3-0.6 percentage points or 76,000-152,000 deaths per year nationally. Analysis based on IV strategy shows that cooking fuel choice has a significant impact on under-five mortality mainly through its effect on neonatal mortality, and the result is robust to a set of alternative specifications with inclusion of various controls and different estimation methods. I also develop a theoretical model of fuel use to demonstrate how this causal effect of cooking fuel choice on infant mortality can come about.

BIOGRAPHICAL SKETCH

Tsenguunjav Byambasuren was born in Ulaanbaatar, a capital city of Mongolia. He completed his bachelor's degree in economics at the University of Finance and Economics in Mongolia in 2013 with honors, majoring in business economics. Tsenguunjav's main areas of research interest are international and development economics, international finance, and trade. Prior to his enrollment at Cornell University, Tsenguunjav has worked as a research economist at the Central Bank of Mongolia in responsible for carrying out fundamental and policy research, conducting macroeconomic modeling and forecasts, and providing monetary and exchange rate policy recommendations.

To my parents, for encouraging all my endeavors in all situations

ACKNOWLEDGEMENTS

First and foremost, I would like to sincerely thank my thesis advisors, Prof. Nancy Chau, Prof. Arnab Basu and Prof. Neha Khanna, for their encouragement, support, and close guidance throughout the project. In particular, I am immensely grateful to Prof. Nancy Chau for her tireless and constant support, wisdom, and confidence in me over the years.

I gratefully acknowledge the funding support from the Government of Mongolia funded Fulbright Scholarship Program for making this journey possible and enabling me to focus on my studies.

Last but not least, I would like to thank my friends, particularly, Avralt-Od, Tulga and Gantushig, and members of the cohort, for the numerous discussions, support, and energy they gave to me during these years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

My thesis has two strands. The first strand empirically estimates the causal effect of cooking fuel choice, a proxy for indoor air quality, on infant mortality in India based on demographic and health survey (DHS) data. Another strand in my thesis explores a theoretical framework of overlapping generations to formulate the association between infant mortality (or likelihood of survival) and cooking fuel choice (which can be seen as a household level public good), supported by my empirical estimation. Thus, overall, my thesis not only provides insights on the causal effect of cooking fuel choice, the number one source of indoor air pollution in developing countries, on under-five mortality in economic literature, but also sheds some light on theoretical foundation of the mortality-fuel choice relationship studies.

There is a rich literature on air pollution-health relationship; however, most of those studies focused on the impacts of ambient (outdoor) air pollution, reaching a consensus that outdoor air pollution damages human health and increases the likelihood of mortality (see, for example, [18], [19], [26], and [71]). However, causality of ambient air quality on health outcomes cannot be completely translated into the effect of indoor air quality on health. Therefore, numbers of epidemiological or medical studies have been grown and found a strong and adverse health impact of indoor air pollution. Although there is a large body of existing literature in epidemiology, they have been questioned due to inadequate controls for health outcomes and lack of convincing identification strategy [28]. Due to this gap, there has been a growing literature in economics that tries to better establish the causality in indoor air quality-health outcome relationship by using randomized control trials (RCTs) and the quasi-experimental methods. However, there is still no study proposed a valid instrumental variable that generates significant variation in households' cooking fuel choice. I address the issue of endogeneity in the first chapter and fill an important gap in the empirical literature of causal impact of indoor air quality on infant mortality, which currently does not account for the potential endogeneity due to simultaneity in this particular relationship. In the search of instrumental variables, I focus on the mechanism. In exploring the mechanisms, I found that tree coverage (or forest cover) and agricultural land ownership drive households' cooking fuel choice.

In the second chapter, I develop a theoretical model explaining household's decision towards cooking fuel choice which can be considered as household-level public good. An analytic framework is motivated by an actual observation from India's National Family Health Survey (NFHS) datasets. Interestingly, this chapter sheds new light on how households or individuals invest in clean cooking fuels and make fertility decisions conditional on the likelihood of survival to maximize their own utility over the infinite horizon. A unique feature of this model is that investment level in clean cooking fuel is endogenously determined within the model in which household cooking fuel decision affects the child mortality. To my knowledge, this is the first time attempting to theoretically model the association between fuel choice and child mortality that I have empirically examined in the first chapter.

CHAPTER 2

THE EFFECT OF COOKING FUEL CHOICE ON INFANT MORTALITY

2.1 Introduction

Indoor air pollution (IAP) is positively associated with infant mortality. IAP is produced mainly by incomplete combustion of polluting fuels used for cooking, heating and lighting and is the single largest environmental health risk factor worldwide [87]. Unfortunately, almost three billion people—41% of the world's population—have been using open fire or simple stoves fueled by dirty fuels (such as kerosene, coal, wood, animal dung, and crop waste) for cooking and as a domestic source of energy for the past three decades [90]. Around 95% of these people are poor and live in low and middle-income countries of Southeast Asia, Western Pacific, and Africa: 80% of the population in China, 82% in India, 87% in Ghana, 95% in Afghanistan, and 95% in Chad rely primarily on polluting cooking fuels [28]. Combination of traditional cooking stoves and polluting fuels generates high levels of hazardous pollutants for health. Each year, close to 4 million people die because of diseases attributable to the indoor air pollution (including heart disease, respiratory disease, stroke and cancer) caused by an inefficient use of polluting fuels for cooking and heating [90].[1] Air pollution is the leading environmental factor for death in India, accounting for about 1.2 million deaths in 2017, nearly 40 percent of which due to poor indoor air quality (Global Burden of Disease 2017). Hence, IAP is a major global risk factor and one of the biggest threats to health in the developing world.

[1] In 2016, IAP from solid fuel use resulted in 3.8 million premature deaths, equivalent to 6.7% of global mortality, greater than the toll due to malaria, tuberculosis and HIV/AIDS combined. Of these deaths, 403,000 were among children under 5 years of age [87, 89].

Since women are mainly responsible for cooking and children spend most of their time with their mothers in developing countries, women and young children (especially, children under five years of age) tend to be more exposed to IAP. Approximately 56% of children under-five years stay with their mother at all times during cooking in India [62, 72]. Thus, the environmental risk from IAP to health are highest among the most vulnerable members of society. For example, about 60% of all premature deaths from IAP globally are among women and children, and one-half of all under-five deaths caused by pneumonia are due to IAP. Hence, under-five mortality incidence has been a subject of interest in the air quality-mortality relationship research.

Exposure to IAP from solid fuel use is responsible for just over 3.0% of the global burden of disease (mortality and morbidity) measured by disability adjusted life years (DALYs)[2] lost worldwide, 5.7% in upper middle-income countries, and 7.8% in low-income countries. IAP from cooking with solid fuels is the biggest cause of DALYs in Southeast Asia and Sub-Saharan Africa, and the third leading cause of DALYs globally [3].[3] Moreover, acute lower respiratory infections (ALRI) including pneumonia is the second dominant cause of deaths in children under five years of age in the world after prematurity, and one-third of ALRI-related deaths are because of poor quality air inside the house [13, 91].

To date, most of the studies on the relationship between air pollution and health focused on the impacts of ambient air pollution in the developed and developing world. In fact, there is a literature reaching a global consensus that

[2]The DALY is the most commonly used measure of national burden of disease and combines the years of life lost due to disability with the years of life lost due to death.

[3]Interestingly, ambient (outdoor) air pollution is the sixth leading cause of DALYs in the Southeast Asia and the ninth leading cause of DALYs in the world [1]. Although emission of ambient air pollution from motor vehicles and industrial facilities has been considered as a bigger threat to public health, IAP has started occupying global attention for reducing mortality and morbidity incidences. In addition, IAP is an important source of ambient air pollution [14].

outdoor air pollution levels significantly damage human health [18, 19, 26, 71]. For example, using shutdowns of industrial factories across different sites in the United States during 1981-82 recession as a random factor for air pollution reductions, the effect of a 1 percent reduction in total suspended particulates (TSPs) has been estimated to be a 0.35 percent decline in infant mortality [19]. The findings on the health impact of air quality vary significantly in developing countries. Some studies found no statistically significant effect of air quality, improved by India's environmental regulations, on infant mortality using a difference-in-differences design on dataset of 572 air pollution monitors in 140 cities [39]. Some studies also suggest that concentration of ambient air pollution level during Indonesia's 1997 wildfires were comparable to IAP concentrations. The fire smoke of Indonesia's wildfires increased the fetal, infant, and child mortality incidences [52]. This air pollution from forest fires in Indonesia affected not only young children but also adults, having an adverse impact on adults' abilities to perform both mentally and physically and other health outcomes [34]. As infants' lungs are highly susceptible to pollutants, they are particularly vulnerable to airborne exposure, and more research is required.

Since studies on the association between ambient air pollution and health outcomes cannot been fully translated to the effects of IAP on health, there has been a growing venue of research solely focus on IAP and health. In particular, nearly 200 publications reported health effects of solid fuel combustion in Chinese households and documented that most of those studies find a strong evidence for adverse health outcomes including chronic obstructive pulmonary disease (COPD), ALRI, asthma, lung cancer, and immune system impairment [93]. One of the earliest works which investigates the health impact of IAP found a high correlation between using a traditional stove and having symptoms of respiratory illness using a linear probability model with variety of controls [27]. The first randomized control trial (RCT) experiments on health effects of IAP are conducted in a city of San Marcos, Guatemala, by [25] and [78]. Using logistic random intercept models, they found that use of improved cooking stoves (*planchas*) has a protective health effect by reducing exposure to IAP and symptoms of headache and sore eyes over an 18-month period. The largest RCT with a 4-year of follow-up was also conducted in rural Orissa, India, providing experimental evidence that improved cookstoves in India did not reduce smoke exposure following the second year of installation, or improve health of recipients and greenhouse gas emissions at all because they were not used regularly and recipients did not invest to maintain them properly [43].

Although IAP, its effect on health and households' economic well-being, and policy measures for reducing IAP are major global issues, the socio-economic (welfare) analysis for this problem is relatively new topic among social scientists. A recent paper evaluated the causal effect of the Indonesian government program, subsidizing households to switch from using kerosene to liquid petroleum gas (LPG), using a quasi-experimental approach [49]. Employing difference-in-differences (DID) method, she found that the program led to a 1.1 percentage point reduction in infant mortality. The effects of IAP should not be limited by health effects, but the impacts on socio-economic outcomes including productivity, school attendance, and labor market should be addressed as well. Existing studies on the impact of cooking fuels and cookstoves has largely focused on health issues, as burning dirty fuels and using traditional cookstoves produce IAP that has negative impact on health. Much more works is needed in this field of study to better understand the causal effect of IAP on economic well-being of households [28].

In this paper, I aim to contribute to this growing literature by estimating the causal impact of indoor air pollution, defined as use of polluting fuels for cooking, on under-five mortality. Specifically, I exploit an environmental feature, forest cover, and agricultural land ownership that had a large effect on cooking fuel types, without any accompanying direct effect on child mortality. This allows us to isolate the effect on household decision for cooking fuels that is driven by mortality cases. Understanding this reverse causality from mortality (health outcome) to cooking fuels is important for several reasons. First, individuals or households may change their choice of fuels used for cooking to prevent from another mortality case due to indoor air pollution if there was a mortality case in their household. This is a reasonable assumption to make because utility maximizing households or individuals, who value their health, are likely to respond to mortality by changing the type of cooking fuel. Second, the existing literature agreed that there is a channel where air pollution adversely affects an individual's long-term earnings through poor health and low productivity [36, 50].[4] The low-income households will be able to afford only cheaper option for cooking fuel or they will be forced to purchase polluting fuel, furthermore, it affects the health outcomes and mortality, as well as household earnings in return [16, 17, 35, 37, 44]. A strong negative effect of air pollution (carbon monoxide–CO) on fourth-grade test scores (math and language skills) in Santiago, Chile, and the 50% increase in CO in Santiago from 1990 to 2005 reduced an individual's lifetime earnings by around US$100 million [9]. Hence, this potential reversal relationship has to be instrumented using an exogenous variable in order to solve a potential endogeneity issue and produce consistent

[4] One of the primary determinants of households' decision about what type of fuel (dirty or clean) to use as a source of energy and/or for cooking is their income or affordability to purchase different types of cooking fuels. The effects of air pollution on labor productivity and human capital complement the impact of air pollution on household income.

estimates. I use two instrumental variables for household fuel choice including forest cover and agricultural land ownership. First, I consider that density of forest across different locations in the country determine the availability or access to firewood, which is often classified as a polluting fuel. Second, I also argue that households which own land for agricultural purpose would more likely to use polluting fuels such as agricultural crop waste, animal dung, and even fuelwood.

Most of the earlier works on the consequences of indoor air pollution, especially, those epidemiological and medical studies, are questioned in terms of reliability of their statistical results due to inadequate controls for health outcomes and lack of convincing identification strategy [28]. Hence, the objective of this study is to evaluate whether polluting cooking fuels affect health outcomes, focusing on infant mortality. I provide the first empirical estimate of the causal effect of IAP, defined by cooking fuels, on under-five mortality by relying on plausibly exogenous variations in IAP introduced by forest cover and agricultural land ownership. Intuitively, forest availability increases the fuelwood collection and consumption in cooking, one of the leading cause of IAP, due to greater access to fuelwood, resulting lower prices of local firewood, and less opportunity costs facing households to collect the fuelwood (less collection time and labor).[5] For instance, in the Brazilian Atlantic forest region, a considerable portion of the rural population still depends on self-harvested firewood from remnant native forests to meet its basic cooking needs [24, 79].[6] The similar

[5]However, some studies found counter-intuitive behavior of households, i.e., distant households tend to consume more firewood [11].

[6]One can argue that firewood consumption causes forest degradation; however, contribution of firewood harvesting to deforestation is negligible compared to other leading factors including climate change, forest fires, agriculture plantations, illegal and unsustainable logging, and mining. A village level data from the East Nepal Hill Region show that fuelwood demand does not significantly affect deforestation like the food supply [5]. In developing countries, population growth and agricultural expansion are the main contributors in the short term, while wood

argument also can be raised for households with agricultural land.

I use a large-scale data set of household survey collected throughout India that recorded the demographic and health information including type of fuels used for cooking from 1992 to 2016. Analysis based on the instrumental variable method shows that a one standard-deviation increase in the use of dirty fuel for cooking increases the risk of under-five mortality by 3.8 percent. I also show robustness of this finding by performing analyses under a variety of specifications with additional controls and fixed effects.

This study makes the following three contributions to the emerging literature on the impact of indoor air pollution on child mortality. First, to my knowledge, this is the first attempt to empirically estimate the causal effect of indoor air pollution on infant mortality while addressing the endogeneity issue in the relationship between cooking fuel choices and health outcome (or mortality) using household survey data and geospatial information. Existing studies are mostly based on epidemiological estimate of the IAP-mortality relationship rather than empirical estimates of the relationship [28]. While the endogeneity issue in the mortality-IAP (or -cooking fuel) relationship has been recognized [76], it has rarely been addressed in empirical settings perhaps due to the challenge in finding a valid instrumental variable. Different from the framework of simple logistic regression in the literature, I use a binary measure of cooking fuel to characterize the indoor air pollution and employ an IV strategy based on forest cover and agricultural land for identification in main analysis.

Second, this study provides the first empirical causal estimates of indoor air

harvesting for fuel and export plays in a role over the long term [2].

pollution which is defined by type of cooking fuels. Previous papers mainly focused on effectiveness of IAP reducing policy and programs (e.g., improved cooking stoves, house construction, and voucher allocation of electrification) on IAP and selected health outcomes and used IAP measures defined by cook stoves [7, 12, 27, 43, 78]. One exception is [49] which estimates the causal effect of indoor air pollution (proxied by a household fuel switching program) on infant mortality using difference-in-differences (DID) estimation strategy. Her study uses a quasi-experimental approach by leveraging the kerosene (polluting fuel) to liquid petroleum gas (LPG–clean fuel) conversion program implemented by the Indonesian government. This paper differs from [49] in terms of cooking fuel coverage (key explanatory variable) and health outcome variable (under-five mortality). I consider a total of 12 types of cooking fuels including kerosene, coal/lignite, charcoal, wood, straw/shrubs/grass, agricultural crop waste, and animal dung as a dirty fuel, and electricity, LPG, natural gas and biogas as a clean fuel. In addition, my empirical analysis covers mortality of four different age-groups including neonatal, post-neonatal, child, and under-five as a health outcome variable. [67] leveraged a complete set of cooking fuels in empirical estimation of IAP on under-five mortality; however, their empirical results are unreliable due to potential omitted variable bias and endogeneity bias. They overestimated the effect of polluting fuel on under-five mortality by providing odds ratio of 1.30, and the upward bias of their estimate could be driven by omitted variables such as number of people in the household, small size of dwelling, birth order or other regional-level demographic and environmental factors that increase IAP concentration and at the same time increase infant mortality.

Third, this analysis adds to the existing literature on health consequences of IAP by utilizing the NFHS (also called as Demographic and Health Survey–DHS)—a widely-accepted gold standard for development research in developing world—datasets which cover a total of 601,509 representative households from all 36 states and 640 districts of India over the last 25 years [51]. Most of the few papers that study the indoor air quality in developing countries largely focused on a rural village of Orissa in India [27, 43], a city of San Marcos in Guatemala [78], and a rural village of La Victoria in the western highlands of Guatemala [12]. A detailed and large-scale dataset collected from this nation-wide household survey, covering both urban and rural areas, allows us to provide reliable indicators and empirical estimates.

The remainder of the paper is organized as follows. Section 2.2 presents the background on IAP and child mortality in India and provides the trend analysis of under-five mortality attributed to the cooking fuel types. Section 2.3 lays out the empirical strategy, and Section 2.4 describes the data and presents descriptive statistics for the sample. Section 2.5 presents model estimation results and set of robustness tests. Section 2.6 concludes.

2.2 Background

In this section, I first discuss India's challenges regarding the IAP induced from widely used solid (polluting) fuels for cooking and its potential harmful effect on health of household members, especially for mothers and children, and their economic well-being. In particular, I focus on the association between IAP and early childhood (under-five) mortality rates.[7] I then present and discuss the

[7]Under-five mortality rate is indispensable gauge of child health and has several advantages as a barometer of child health and child well-being in general. First, it measures an 'outcome'

trends of India's under-five mortality rate in relation to type of cooking fuels.

2.2.1 Indoor Air Pollution and Infant Mortality in India

India is the second-most populous country (with over 1.3 billion people) and seventh-largest country by area in the world, and tenth-biggest contributor to the global GDP (1.9% of the world nominal GDP on average from 1960 to 2017). However, 72.3% of households (more than 90% of the rural population and 31% of the urban population) in India still use solid fuels as a primary source of energy and for cooking. The United States Environmental Protection Agency sets standards for PM_{10}[8] concentrations to 50 μg/m^3 based on an annual average, and 150 μg/m^3 based on a 24-hour average. Additionally, the EPA states that these levels should not exceed more than once per year. However, 24-hour average of PM_{10} concentration in solid fuel firing households in India sometimes exceeds 2,000 μg/m^3 [77].[9]

According to World Health Organization (WHO), 3.5% of the total burden of disease in India has been caused by IAP [85], while 20% of deaths among children aged under-five is attributed to the use of solid fuels [8, 82]. Using data from Global Burden of Disease 2017, [6] estimated that 1.2 million deaths

of the development process rather than an 'input', such as per capita calorie availability or the number of doctors per 1,000 population – all of which are means to an end. Second, the under-five mortality rate is known to be the result of a wide range of inputs such as the nutritional status and the health knowledge of mothers, the level of immunization and oral rehydration therapy, the availability of maternal and child health services, income and food availability in the family, the availability of safe drinking water and basic sanitation, and the overall safety of the child's environment. Third, it presents much more accurate picture of the health of children (and of society as a whole) than per capita gross national income [81].

[8] PM_{10} describes inhalable particles with diameters that are 10 micrometers and smaller.

[9] For India, [64] and [75] found that much higher concentration of air pollution (20,000 μg/m^3) is exposed near the cooking location, and level of concentration substantially decreases as moving away from kitchen.

in India in 2017, which were 12.5% of the total deaths, were attributable to air pollution, including 0.7 million to ambient $PM_{2.5}$ and 0.5 million to IAP. In addition, [67] estimated the negative effect of IAP on under-five mortality using India's National Family Health Survey (NFHS) datasets during the years 1992–2006. Several other attempts have been made to quantify the cost of IAP in terms of productivity, for example, [77] estimated that the annual health burden for India from IAP is 1.6-2.0 billion days of work lost (number of sick days due to the diseases caused by IAP). Besides of reducing adult productivity, poor health attributable to IAP also affects the schooling and productivity of children. For example, large portion of absence from schooling in rural areas of India is because of poor health [29].

Due to perceived health threats from polluting fuels, Indian authorities and non-governmental organizations (NGOs), like many other governments, NGOs and international organizations, have been implementing policy strategies and programs for reducing IAP. For example, subsidizing cleaner fuel technologies, distributing "improved cooking stoves", and convincing households to improve ventilation system within the household are common interventions. Among these policy strategies, the improved cooking stove has become the most popular policy prescription for reducing IAP. The Government of India implemented the second largest program in the world to limit emission of smoke within households and distributed around 33 million biomass-based improved stoves in rural areas during 1984-2000 through its National Biomass Cookstoves Programme, while China leads the rank having over 35 millions of improved stoves distributed. The list of the countries devised this strategy continues with countries in Sub-Saharan Africa such as Ethiopia and Kenya, installing 1.5 million improved stoves for more than 10 years [28]. It is well established that the stoves promoted by National Programme on Improved Chulha (NPIC) in India with fuel-use efficiency of 20-35% reduced the fuel consumption and the time and effort that rural women put into collecting fuel per meal by half. Aiming to save almost 20 million tons of fuelwood per year, this government initiative had not been a failure. However, the effectiveness in reducing IAP and health benefits of the programme were far below the expectations. Many studies suggest that the "improved" cooking stoves had even a hazardous impact on health due to an inefficient use [42, 57].

2.2.2 Mortality Trends in India

The overall under-five mortality incidence proportion in India has been declining since 1992 until 2016. The total mortality rate for under-five children decreased from 11.6% (of which 8.5% for those using polluting fuel for cooking and 3.1% for those using clean fuel for cooking) in 1992 to 7.4% (of which 4.7% for those using polluting fuel for cooking and 2.7% for those using clean fuel for cooking) in 2016. Here, we can observe that under-five cumulative mortality incidence for those using polluting fuel for cooking has been dramatically (by about 45% for the last 25 years) decreasing, while under-five mortality rate for those using clean fuel for cooking has been leveled off at around 3% per year (Figure 2.1). Note that under-five mortality incidence proportion for those using polluting fuel for cooking is significantly higher than that for those using clean cooking fuel. This result is quite intuitive and indicates that risk of child's death is much higher in the household which uses polluting fuel for cooking. The gap between the mortality rates for these two types of households has been continuously shrinking; however, under-five mortality is still close to 100% higher in households which use solid fuels for cooking compared to those use clean cooking fuels. Despite the decreasing trend, the improvement in India's mortality incidences for the last decade might be unsatisfactory.

If we look at mortality rates for three age-groups, we can observe that neonatal (first 28 days of life) mortality rate is the highest among those three age groups, followed by post-neonatal (period between approximately the first month after birth and end of the first year of life) mortality and then child mortality. This is due to the fact that younger children have weaker immune system and higher risks (probability) of mortality compared to the older ones. Decreasing trends are also observed for each age group, where the neonatal mortality rate declined from 4.4% in 1992 to 3.1% in 2016, post-neonatal mortality rate from 2.6% in 1992 to 1.1% in 2016, and child mortality rate from 1.5% in 1992 to 0.5% in 2016 for those using polluting fuels for cooking (Figure 2.1).

2.3 Empirical Strategy

In this section, I first describe the empirical specification for the relationship between IAP and child mortality. I then discuss the empirical challenges in estimating the causal effect of IAP on under-five mortality.

2.3.1 Indoor Air Pollution and Infant Mortality

To empirically examine the causal effect of IAP on mortality of children under five years of age, I specify the following relationship between IAP and under-five mortality:

$$Y_{ihvdst} = \alpha + \beta D_{hvdst} + \boldsymbol{X}_{hvdst}\boldsymbol{\gamma} + \boldsymbol{M}_{jhvdst}\boldsymbol{\lambda} + \boldsymbol{W}_{ihvdst}\boldsymbol{\delta} + \tag{2.1}$$

$$+\mu_t + \eta_{ds} + (\eta_{ds} \times \mu_t) + \varepsilon_{ihvdst}$$

where Y_{ihvdst} is one of the four binary variables for under-five mortality taking the value 1 if the death occurred during the considered age-periods (neonatal, post-neonatal, child, and under-five) and 0 if the child survived during the age-period for child i, in household h, in village v, in district d, in state s, in survey year t. The key regressor is a binary variable for solid fuel use (D_{hvdst}) in household h, in village v, in district d of state s, in year t as defined above. The vectors $\boldsymbol{X}_{hvdst}$, $\boldsymbol{M}_{jhvdst}$, and $\boldsymbol{W}_{ihvdst}$ are respectively composed of household (h) characteristics including place of residence, household wealth index, number of household members, place where food is cooked and type of house, mother (j) characteristics including mother's age and mother's education, and child (i) characteristics including gender of the child and breastfeeding status. The error term, ε_{ihvdst}, is the unobserved, time-varying, and child-specific factors.

The district fixed effects, η_{ds}, control for all permanent unobserved determinants of mortality across districts, while the inclusion of the year fixed effects for year of survey, μ_t, nonparametrically adjust for national trends in under-five mortality, which is important in light of the time patterns observed in Figure 2.1. The relationship could also vary across districts (use of polluting fuels differs across different regions) and across time (trend of using solid fuels, IAP, changes over the period). To control for possible unobserved spatial differences in IAP at different periods, I interact the time fixed effect with districts. In other words, to control for other confounding factors that may vary across time but are not adequately controlled by the time fixed effects, I include a district-specific time trends, $\eta_{ds} \times \mu_t$, to allow the unobserved time trend to vary across districts. Controlling for District×Time fixed effects allows us to estimate the effect of region-specific characteristics varying over time, which can be seen as regional (or neighborhood) differences such as culture, weather conditions, environmental features, and local-level policies or programs on cooking fuels.

2.3.2 Identification

The key identification challenge is the potential endogeneity of the IAP resulting from non-random use of polluting fuels. In the empirical literature on air pollution and its health consequences, it is assumed that IAP affects mortality and other human health outcomes but not vice versa. In practice, IAP and choice of fuel types for cooking can be affected by the mortality, morbidity, and other health outcomes. The existing literature on the association between IAP and health, for example, [28] documented the evidences on the potential effects of IAP on health and productivity (in overall, economic well-being). The previous works in the field generally reached a consensus that there is a channel where IAP affects an individual's long-term earnings. However, low-income households can only afford the cheaper option for cooking fuel or they will be forced to purchase polluting fuel,[10] furthermore, it affects the health and earnings. This simultaneity would give rise to endogeneity in IAP in Equation (2.1). Hence, one must cut this backward channel from health outcome to IAP by specifying an instrumental variable (IV) to provide consistent estimates.

[10]It should be obvious that the households' decision to use polluting (solid) or clean fuel as a source of energy and/or for cooking is determined by their earnings (affordability to purchase different types of cooking fuel). I use household wealth index as a proxy for earnings/income of the household.

The first stage of the IV strategy specifies IAP as a function of instrumental variables (IVs) and other controls.

$$D_{hvdst} = \boldsymbol{Z}_{dst}\boldsymbol{\pi} + \boldsymbol{X}_{hvdst}\boldsymbol{\gamma} + \boldsymbol{M}_{jhvdst}\boldsymbol{\lambda} + \boldsymbol{W}_{ihvdst}\boldsymbol{\delta} + \\ + \mu_t + \eta_s + (\eta_s \times \mu_t) + \xi_{hvdst} \tag{2.2}$$

where $\boldsymbol{Z}_{dst}$ is district-specific characteristics and η_s is the state fixed effects. Notice that I use district-specific characteristics even though the village-level information is available in the Census data. This is due to random PSU point (or village/city block)[11] displacement in the NFHS GPS data, which limits us to correctly match the PSUs with Census locations only at country, state, and district-levels.[12] In other words, I am not able to correctly match the NFHS dataset with Census data at sub-district and village levels as the maximum displacement buffers for particular cluster points overlay with level 3 administrative (sub-district) boundaries [69]. Figure 2.3 shows the displacement strategy of PSU points (GPS information of respondent locations) in NFHS-4 and reason why I am unable to correctly identify actual sub-districts and villages where the NFHS survey respondents live in. Although the PSU point displacement is randomly conducted, it would affect my empirical estimation because I am matching NFHS data with Census data by location to merge the potential IVs from Census data into the NFHS dataset.

[11]Figure 2.2 depicts how the PSUs in India's NFHS-4 (2015–16) are distributed across the country.

[12]According to the description of the NFHS GPS data provided by the DHS Program, the displacement is restricted so that the PSU points stay within the country, the NFHS survey region (state), and district area. Therefore, the displaced cluster's coordinates are located within the same country, state, and district areas as the undisplaced cluster. This random error can substantively affect analysis results, where analysis questions look at small geographic areas including sub-districts and villages/city blocks.

2.4 Data

I use detailed household and individual-level demographic and health data and geospatial information of forest cover from India to analyze the positive relationship between IAP (dirty cooking fuels) and child mortality. The features of the core data sets that are most relevant for this analysis are described below. My empirical analysis is based on two datasets. The first data set (nearly 0.4 million observations) is nationally-representative data from India's National Family Health Survey (NFHS) datasets. To date, four rounds of the survey have been conducted since the first survey in 1992–93.[13] The analysis relies heavily on three different rounds of this household survey data, between 1992 and 2016: NFHS-1 (1992–93), NFHS-2 (1998–99), and NFHS-4 (2015–16).[14]

The NFHS collects various individual-level data on mortality incidence and other socio-economic characteristics of every member in the sample household. Additionally, NFHS data also contains a variety of household-level survey data related to wealth, housing, and residence. Importantly, NFHS data includes information on type of cooking fuel allowing us to approximate indoor air qual-

[13] The NFHS, a nationwide survey, is conducted under the supervision of the Ministry of Health and Family Welfare, coordinated by the International Institute for Population Sciences, Mumbai, and implemented by a group of survey organizations and Population Research Centres. Technical assistance for NFHS is provided by the USAID-supported DHS (Demographic and Health Surveys) Program at ICF International, USA with the major financial support from the United States Agency for International Development and Ministry of Health and Family Welfare, Government of India. While the first three NFHS survey datasets cover all 29 states of India, which includes more than 99% of India's population, the most recent NFHS data for the years 2015–16 (NFHS-4), fourth survey in the NFHS series, adds all six union territories for the first time. It is worth noting that I define state by including union territories, i.e., I treat union territories as states (UTs = States). The NFHS-4 also provides vital estimates of most demographic and health indicators at the district level for all 640 districts in the country (as per the 2011 Census).

[14] This analysis was unable to use the NFHS-3 (2005–06) in my empirical estimation due to absence of district identifiers (either names or codes) in the questionnaire of this particular round for the purpose of confidentiality of HIV testing. In fact, I had to drop 50,750 observations from NFHS-3 since I include district fixed effects while there is no district names.

ity at the household level. A total of 1,003,880 ever-married women of reproductive age[15] (317,250 from urban and 686,630 from rural areas) were included in three datasets (NFHS-1, NFHS-2, and NFHS-4) that I analyzed in this paper. This study is based on pooled dataset of 421,709 singleton (twins are excluded) live-born children, of whom 22,268 died in the 5-years prior to the survey. This dataset also contains household's agricultural land ownership status (or whether household owns land for agricultural purpose).

Second dataset, 2011 Census of India, includes land use information at the village and city block level. The data set contains the total surface area of the land of each geographical region and land covered by forests in hectares (ha). It is important to include the forest condition as it could serve as a valid instrumental variable for the key explanatory variable, type of cooking fuel.

2.4.1 Under-Five Mortality

Under-five mortality rates are an appealing measure to be used for estimating the effect of indoor air pollution for at least two reasons. First, children under five years tend to spend most of their time at home alongside their mothers most of whom are responsible for cooking, and so under-five children are largely exposed to indoor air pollution. Second, earlier years of life are especially vulnerable periods, and so losses of life expectancy are likely to be large. I examine four outcome variables for mortality during four different age-groups. Primarily, the outcome variable in this study is the under-five mortality. In addition, to examine how the relationship changes for early ages of children, the analysis for

[15] A reproductive age refers to 15–49 years in this paper and ever-married women aged less than 15 are excluded from the sample. All the women interviewed in the survey were ever-married and only 271 of them were aged less than 15 years, i.e., women aged 13 and 14 years.

under-five mortality was conducted for three preceding age groups: neonatal, post-neonatal, and child mortality, using the following definitions:

- *Neonatal mortality:* The number of deaths during the first 28 days of life (0–28 days), defined as (Number of neonatal deaths/Total number of live births);
- *Post-neonatal mortality:* The number of deaths between one month and the first birthday (1–11 months), defined as (Number of post-neonatal deaths/Total number of live births);
- *Child mortality:* The number of deaths between exact ages one and five (12–59 months), defined as (Number of child deaths/Total number of live births).

The mortality rates (cumulative mortality incidence, %) calculated by the ratio defined above have been used for the trend analysis presented in Figure 2.1, while the outcome variables in the regression analysis are considered as dichotomous, i.e., age at death takes a value of 1 when the death occurred during any of these three periods of age and takes a value of 0 when the child survived during the age-period.

2.4.2 Cooking Fuel Choice and Other Controls

The key explanatory variable is type of fuel used for cooking in the household, a proxy for indoor air pollution, and it serves as a main exposure variable. A total of 12 types of cooking fuel are reported in NFHS datasets, and I classify these fuels into two groups (clean and polluting) depending on exposure to cooking smoke by following the fuel energy literature. The clean fuel includes electricity, liquid petroleum gas (LPG), natural gas and biogas, while polluting fuel includes kerosene, coal/lignite, charcoal, wood, straw/shrubs/grass, agricultural crop waste, and animal dung. Note that no household recorded as using more than one types of cooking fuels in the survey.

In addition to the main exposure variable, several other determinants of under-five mortality are collected from the NFHS datasets. Place of residence (urban or rural), household wealth index (high income, middle income, or low income),[16] number of household members,[17] mother's education (secondary or higher, primary, or no education), and type of house (pucca, semi-pucca, or kachha) are included as potential socio-economic factors [8, 31, 32, 53, 66, 67, 68, 73, 74, 80, 92]. Mother's age (<20, 20–29, 30–39, and 40–49 years) and gender of the child (female or male) are also considered as potential confounders of the association between IAP and under-five mortality.

Breastfeeding status of mother (ever breastfed or never breastfed) and place

[16]The household wealth index was constructed using principal components analysis, with weights for the wealth index calculated by giving scores to the asset variables such as ownership of transport, durable goods, and facilities in the household. "Low income" referred to the bottom 40% of households, "middle income" referred to the middle 40% of households, and "high income" referred to the top 20% of households [33].

[17]It is the total number of members living together in a household and not necessarily a family size. There are households with members up to 46 people in the household and households covered in the NFHS have about 7 members on average. In India, wealthier households usually have more dependents living at home and correlation coefficient of 0.09 between household wealth and size confirms this observation (if I calculate the correlation of household size with each income group, $\rho_{size}^{high} = 0.07$, $\rho_{size}^{middle} = 0.04$, and $\rho_{size}^{low} = -0.09$). Hence, one may expect household size effect to dominate the effect of household wealth in explaining choice of cooking technology. Figure 2.4 depicts a scatter plot between mean fraction of polluting fuel use and number of household members and it shows that there is a strong correlation between household size and fuel choice. In other words, use of polluting fuels tends to increase as family size get larger. Gas stove limits the volume of food that can be cooked because the size of the stovetop is small. Wood burning furnaces can be built to accommodate larger utensils which would be preferred by households with number of members. A right-skewed histogram of household size illustrated in Figure 2.5 suggests that households with less than about 25 members are quite prevalent in the data while households with more than 25 members could be considered as outliers.

where food is cooked (in the same room as they live inside the house, in a separate room as kitchen inside the house, in a separate building, or outside)[18] are also considered a priori factors that may indicate different levels of exposure to polluting fuels. No separate kitchen used for cooking inside the house as an indicator of proximity to polluting fuel use has also been presented to be an significant factor associated with high exposure to IAP [23, 41, 54, 55, 56, 66, 67, 73]. Additionally, breastfeeding has been shown to be a protective factor for under-five mortality, generally in neonatal and infancy period [4, 10, 21, 32, 45, 92] which may reduce the greater risk of exposure associated with IAP. Hence, I seek to determine whether the magnitude of the association between IAP and under-five mortality differ by past breastfeeding status.

In terms of exposure to IAP, this is the first time controlling for whether food is cooked inside,[19] in a separate building, or outside in the economic literature of IAP due to availability of indicator for place where food is cooked in NFHS-4 (2015–16) for the first time. Earlier rounds of the NFHS survey only included a variable indicating whether the household has a separate room as kitchen inside the house. I consider that controlling for place of cooking (inside/outside/separate building) combined with indicator for separate kitchen inside the house is important in estimating the causal impact of cooking fuel

[18] In the NFHS questionnaire, the question asking whether the household has separate room as kitchen captures only cooking inside the house. So, the variable for location of kitchen is not relevant to outdoor cooking. Another variable for indicating if the household cooks inside the house, in a separate building, or outdoors is only available in NFHS-4 (2015–16). Therefore, if I control this variable in the regression, I am forced to use only the last round of NFHS survey. A separate room for cooking and/or as kitchen may not make much of a difference as compared to cooking inside and/or in the same room as they live in because ventilation within house is not that developed in India, especially, in rural area, and smoke permeates everywhere if cooking with wood or coal.

[19] I separate cooking inside the house into two groups: in a separate room as kitchen inside the house or in same room as they live inside the house based on variable indicating an existence of separate kitchen, i.e., question asking whether the household has a separate room used as kitchen in house.

choice on under-five mortality. This is because without this control, when the household does not have a separate kitchen in house (i.e., they cook in the same room as they live in), I treat that as they cooked outside or in a separate building (or No kitchen = Outside/Separate building). The air quality impact on cooking in the same room is different from cooking outside or in another unit.

2.4.3 Instruments for Cooking Fuel Choice

To account for endogeneity, I specify the location of forests for identification by arguing that this environmental condition generates exogenous variation in cooking fuel choice (or IAP) that in turn affects under-five mortality. The lack of data on prices of firewood and LPG at district and/or village level makes us unable to introduce significant variations in cooking fuel choice by controlling for main fuel prices. Note that the main fuels for cooking among nearly 891,600 households covered in all four rounds of India's NFHS are wood and LPG, accounting for 82.5% of total households covered in the survey over the period 1992–2016 (Figure 2.6). Hence, I use forest cover as a proxy for relative price (cost) of firewood. The location of forests satisfies the exogeneity assumption because it changes exogenously regardless of the children mortality. Therefore, the forest location is uncorrelated with unobserved, time-varying, and child-specific shocks that affect under-five mortality.

The forest cover is relevant and generates meaningful variation in cooking fuel choice (or IAP). First, wood is the most widely-used fuel for cooking in India. Figure 2.6 shows the percentage of cooking fuels in India. Given that one-half of the Indian households covered in four rounds of the NFHS rely on wood as a fuel for cooking, the forest cover in India unintentionally allows more households to consume a fuelwood (polluting) for cooking. Second, the availability of wood (solid or polluting fuel) significantly influences its consumption. The population living near to the forest resources has higher per capita consumption of wood compared to those living farther from forests. Households living in villages with forests use fuelwood twice as much as that of households in villages without forests [70]. Therefore, I assume that households in the forested regions tend to rely on fuelwood as their source of energy and for cooking and have less motivation to use clean fuels even they can afford the expensive clean fuels (relevance assumption). Figures 2.7 illustrates the district-wise forest cover in states of India.[20] Share of households using solid fuels for cooking in three of the largest five forest cover states (89% in Odisha, 85% in Chhattisgarh, and 82% in Madhya Pradesh) are substantially larger than the country average, 77%, confirming that location of forests affects IAP from burning solid fuels for cooking. Furthermore, under-five mortality rates in these three states (6.0% in Odisha, 5.8% in Chhattisgarh, and 6.6% in Madhya Pradesh) are persistently higher than the country average, 5.3% (Table 2.2). This geographic variable hence induces plausibly exogenous variation in IAP that is not correlated with the unobserved, time-varying, and child-specific shocks to under-five mortality.

The well-known concept of energy ladder[21] describes that wood is the

[20]According to the 2017 State of Forest Report, a biennial publication by the Ministry of Environment, Forest and Climate Change, India's forest covers 708, 273 square kilometers occupying 21.54% of the total land area of the country. While forest and tree cover is concentrated in the central region, very dense forests are located in the northeast region. The satellite data from the LISS-3 sensor suggests that Madhya Pradesh (77, 414 km^2) is the largest forest cover state in the country followed by Arunachal Pradesh (66, 964 km^2), Chhattisgarh (55, 547 km^2), Odisha (51, 345 km^2), and Maharashtra (50, 682 km^2). Thus these five states account for 42.63% of total forests.

[21]The process of climbing the energy ladder illustrates transitions in fuel choice as household income or economic development at large improves. Poor households would be at the bottom

second-most polluting fuel, after crop waste and/or dung being the dirtiest solid fuels for cooking [28, 88]. Although carbon monoxide (CO) and particulate matter ($PM_{2.5}$) emissions in terms of micrograms per cubic meter ($\mu g/m^3$) from coal/charcoal is higher than that from fuelwood, wood is seen as a lower quality fuel due to its thermal inefficiency. Thus, wood generally appears lower on the energy ladder, but [48] argue that coal is dirtier than wood. Another reason why wood plays important role in cooking in developing countries is that it is the cooking fuel used by the widest range of households in terms of income-level, i.e., very low-income, low-income and middle-income households use wood for cooking. Consistent with this observation, around 72% of Indian population relied on solid fuels in 2007, of which 58% used wood for cooking [60], suggesting that wood is the most widely used solid fuel and the biggest source of indoor air pollution in India.

Hence, I collected additional data on sociodemographic and environmental characteristics at the village and city block-level. Environmental data come from two sources. First, I obtained village-level data on land covered by forests (in hectares) from 2011 Census of India. Since availability of fuelwood or access to fuelwood should be expressed in relative term rather than in absolute term of area (ha), I defined forest cover as a land covered by forests per capita (ha/person) and percentage of total geographical area of the village (% of land area). The village-level data on population and area of village (in hectares) come from the 2011 Census of India. This measure of forest cover defined above was used as an instrumental variable for cooking fuel. Second, to examine spatial and temporal variation in forest cover, I collected district-level data on forest

of the energy ladder and use less costly, inefficient, and more polluting fuels such as firewood, agricultural waste, and animal waste. At the next steps of the energy ladder, households rely on fuels including kerosene, coal, and charcoal. In the last steps, fuels households switch to fuels such as biogas, LPG, and electricity [28, 48, 88].

cover from the Planning Commission of India (reformed as the National Institution for Transforming India–NITI Aayog in 2015) for three years including 2007, 2011, and 2013. In this dataset, forest cover refers to all lands more than one hectare in area, with a tree canopy density of more than 10 percent irrespective of ownership and legal status. It also includes orchards, bamboo and palm. Figure 2.7 plots India's district-wise forest cover as in 2007.

Table 2.1 presents summary statistics on cooking fuels, infant mortality, and other demographic indicators used in the regression analysis. The data suggest that under-five mortality rate during the period that I analyzed were 5.3% and infant mortality rate increases as age of child decreases. Majority (76.9%) of the households covered in the survey use polluting fuels, while the remaining portion of the households use clean (electricity, LPG, and biogas) fuels for cooking. Three-fourth of the individuals (or children) included in the analysis are from rural region, whereas the remaining one-fourth live in urban area. So I am analyzing individuals reside in both urban and rural regions, although the mainstream of the IAP literature mainly focuses on rural households. Since mothers in the rural area are not likely to have a job, about three-fourth (73.2%) of the mothers did not have job. In addition, most (67.9%) of the mothers with children aged under five are intuitively mothers who are 20-29 years old since mothers who are younger than 20 years old and older than 29 years old are less likely to have children aged under five. In terms of other socio-economic characteristics including household income, mother's education, gender of child, location of kitchen, and type of house, the individuals included in the analysis are well balanced.

Tables 2.2–2.4 present summary statistics for the four outcome variables (infant mortality for different age-groups) and key explanatory variable (type of cooking fuel) by geographic region, age of the household head, and gender of the household head. The data indicates that infant mortality rate and fuel choices significantly vary across regions throughout the country (Table 2.2). Moreover, infant mortality and fuel choices are relatively constant across different age groups (Table 2.3) and gender (Table 2.4) of the household head.

2.5 Results

In this section, I first present the estimated average marginal effects[22] of cooking fuel choice (indoor air pollution–IAP) on infant mortality using the multivariate probit and the IV (2SLS) regressions. I then discuss the implication of baseline estimation results and set of robustness tests.

2.5.1 Regression Results

Table 2.5 presents the results of Equation (2.1) for under-five mortality under five different specifications where I add more control variables successively. All model specifications are estimated on the full sample using the multivariate pro-

[22]Marginal effects are generally computed using two methods including average marginal effects (AME) and marginal effects at the means (MEM). MEM is calculated by setting the values of all covariates to their means within the sample. To obtain the AME, the marginal effect is first calculated for each individual with their observed levels of covariates, and these values are then averaged across all individuals. Since independent variables including key regressor, fuel choice, are binary variables, the average marginal effects measure *discrete change* or how do predicted probabilities (infant mortality) change as the binary independent variables change from 0 to 1. For probit regression, the average marginal effects are calculated by

$$\text{Marginal Effects for } \mathbf{X} = \Phi(\mathbf{X})\boldsymbol{\beta}$$

where Φ is the probability density function for a standardized normal variable.

bit regression. The average marginal effect of the key regressor, use of polluting fuel for cooking, ranges from 3.3 to 0.8 percentage points in the five regressions with the last two regressions producing the smallest estimates. Since the coefficient estimates and calculated marginal effects on polluting fuel use is consistently greater than zero and statistically significant at 1 percent level for each specification, I suggest that indoor air pollution (use of polluting fuels for cooking) has been contributing to the mortality risk among children aged under-five in India. I consider the last regression as my preferred specification because the inclusion of interaction of district dummies with year fixed effects controls for time varying spatial factors such as district attributes and local characteristics that could affect both under-five mortality and fuel choice.

To examine the effect of cooking fuel choice on infant mortality in more detail, I consider three alternative scenarios for mortality of different age groups: neonatal, post-neonatal, and child. The overall under-five mortality fully contains these three preceding age groups. Table 2.6 presents the results for the first scenario of child mortality. The average marginal effect of IAP on mortality decreases significantly from 0.8 to 0.2 percentage points, as well as the effect of other confounders changes dramatically. The results for the post-neonatal mortality are presented in Table 2.7. The average marginal effect of IAP on post-neonatal mortality is estimated at 0.1 percentage point; however, it is not statistically significant.

Table 2.8 shows the results for the third alternative scenario of neonatal mortality. Compared to other two relatively older age groups, average marginal effect of IAP on neonatal mortality is estimated at 0.6 percentage point, the largest estimate among these three alternative age groups. The comparison is intuitive because with a younger age of child, a greater the effect of IAP. In other words, one would expect that the youngest age should have the largest coefficient estimate because the neonatal period is the most vulnerable time for a child's survival. Except for under-five mortality, the regression results suggest that the harmful effect of IAP on infant mortality increases as child becomes younger, which is consistent with the existing child's age-risk of dying (or -child's vulnerability) argument. Overall, the comparison between the baseline results in Table 2.5 and those under the three alternative outcomes in Tables 2.6–2.8 suggest that the key results are robust to the range of plausible age differences of children mortality from the literature. An important implication of this finding is that the harmful effect of IAP can be reduced by improving the care for infants in order to increase the immunity.

The average marginal effects of the other variables are all intuitively signed and are consistent with infant mortality literature. The risk of mortality in mothers who had never breastfed is the highest compared to other confounders, which is in line with the previous findings. While the infant mortality is higher for teenage mothers, the risk of child mortality decreases for mothers with ages between 30 and 39 years old. In addition, the regression results suggest that improving mother's education is an important factor for reducing under-five mortality incidence. Infant mortality is also higher in households of middle and low income, without separate kitchen inside the house, and live in houses with lower standard. I also include an additional variable of exposure to IAP (whether food is cooked inside/outside/in separate building), which is collected for the first time in NFHS-4 (2015–16). Simple probit regressions show that cooking outside and in a separate building essentially have zero effects on infant mortality (Column (5) of Tables 2.5–2.8).

I compare results from simple multivariate probit regressions with those from [67] which use similar NFHS or DHS datasets for different years, i.e., NFHS-1 (1992–93), NFHS-2 (1998-99) and NFHS-3 (2005–06), to answer the same question, but without any discussion on endogeneity or potential instrumental variables for cooking fuel choice. In doing so, their paper has been replicated and arrived at sufficiently close results (Column (2) of Table 2.9). The results (or odds ratio) of [67] are copied in Column (1) of Table 2.9, while my replication results and calculated average marginal effects are shown in Column (2) of the same table. Since I utilize the most recent round of NFHS, or NFHS-4 (2015–16), I estimated simple logistic regression with exactly the same specification as [67] to check the sensitivity of their results with respect to additional data. Column (3) and (4) present the estimated odds ratios and corresponding marginal effects on sub-sample of only NFHS-4 (2015–16) and a complete sample between 1992–2016 (NFHS-1-4), respectively. Compared to my own calculated average marginal effects for simple multivariate probit regressions, the replicated (or [67]) average marginal effects of polluting fuel use on infant mortality are persistently higher, specifically, approximately two-folds decrease in my estimates. In other words, my specification has lowered the average marginal effects of polluting fuel use on infant mortality. The marginal impacts of type of cooking fuel changes from clean to polluting (from 0 to 1) on infant mortality computed from my non-IV regressions are much smaller (in magnitude) than that found by previous studies, suggesting that the literature overestimated the association between cooking fuel choice and infant mortality.[23]

[23]One could say that I better estimating these simple probit models at PSU (village/city block) level by controlling for PSU fixed effects instead of district fixed effects; however, I did not have enough power to estimate the coefficients at the cluster level because there is nearly 30,000 PSUs or clusters.

I address the endogeneity of cooking fuel choice using IV strategy in Table 2.10, which presents the estimates from the 2SLS regressions where interactions between district dummies and year-of-survey dummies are used as IVs. For the first-stage regression, the coefficient estimates on the interaction terms (IVs) tend to be statistically significant because the joint F-statistic on the excluded instruments is more than 35 (Column (1) of Table 2.10). The interactions of district dummies with year-of-interview fixed effects therefore provides significant variations in fuel choice that I can leverage to identify a causal effect of IAP on infant mortality. Columns (2)–(5) of Table 2.10 present the IV (2SLS) regression results for four different outcome variables and the similar specification. The coefficient estimates on polluting fuel for cooking from the IV regressions of under-five mortality and neonatal mortality are positive and statistically significant, ranging from 0.035 to 0.038. This IV strategy indicate that district-level characteristic varying over time generates variations in fuel choice. However, when I use district-by-year fixed effects as IVs for fuel choice, I am not able to know what exactly has been generating variations in IAP. Hence, I explore forest cover and agricultural land ownership as respectively region and household-specific characteristics, which create exogenous variations in fuel choice of the households and serve as IVs for our endogenous variable.[24]

Before discussing the results found by using forest cover and agricultural land ownership as IVs for the fuel choice, I first present evidence on how they relate to individual's choice of fuel types used for cooking, the relevance assumption for a valid IV. Figure 2.8 illustrates the positive relationship between use of polluting fuels for cooking and forest cover at both state and district levels.

[24]The bivariate correlations of under-five mortality with agricultural land ownership and forest cover are 0.0038 and −0.0137 suggest that IVs are hardly correlated with the outcome variable and hence should not be directly included in Equation (2.1). Thus, I argue that exclusion restriction for IVs is satisfied.

It shows that households tend to use polluting fuels more as the region where the households live have more forests or trees. Moreover, Figure 2.9 plots the positive association between use of polluting fuels for cooking and agricultural land ownership at state and district levels. The correlation coefficients of forest cover and agricultural land ownership with mean fraction of polluting fuel use for cooking at the state-level (0.11 and 0.63, respectively) and the relationships plotted in Figure 2.8 and Figure 2.9 verify that forest cover and agricultural land ownership are relevant for cooking fuel choice.

Table 2.11 presents the first-stage results for Equation (2.2) and the five specifications where demographic controls and different sets of fixed effects are successively included. The IV includes the forest cover measured at the district-level. The forest cover is a continuous variable indicating how many percentages of geographical area is covered by forests as of 2011. The variable has a positive and statistically significant effect on fuel choice, consistent with the fact that districts having wide area of forests account for more likelihood to use polluting fuel for cooking. Therefore, there are more households using solid fuels when their surrounding region has more trees. For the preferred specification shown in Column (5) of Table 2.11, the joint F-statistic on the excluded instrument is above 45. The forest cover therefore provides exogenous variations in fuel choice that I can leverage to identify a causal effect of fuel choice on infant mortality.

Moreover, Column (1) of Table 2.12 reports the first-stage results when indicator variable for household's agricultural land ownership is included as IV. Agricultural land ownership is a dummy variable indicating if household owns land for agricultural purpose in a given year, and it has a positive and statistically significant impact on fuel choice.[25] Since agriculture and its related sectors such as animal husbandry, forestry and fisheries account for 15–20% of India's gross domestic product (GDP) each year (Central Statistics Office, Ministry of Statistics & Programme Implementation, Government of India), agricultural households are likely to consume their own agricultural crop waste and animal dung as cooking fuels, which are classified as polluting fuel. Thus, the result is in line with this observation and suggests that agricultural land ownership also generates a plausible variation in fuel choice. Columns (2)–(5) of Table 2.12 present the estimates from the IV (2SLS) regressions for four different age groups. The coefficient estimates on polluting fuel for cooking from the IV regressions of under-five and neonatal mortality are positive and statistically significant, ranging from 0.033 to 0.040.

Column (1) of Table 2.13 presents the first-stage results when district-wise forest cover and an indicator variable for household's agricultural land ownership are included as IVs. The both instruments have positive and statistically significant association with polluting fuel choice. For this specification, the joint F-statistic on the excluded instruments is large enough to suggest that these two IVs provide exogenous variations in fuel choice that I can leverage to identify a causal effect of fuel choice on infant mortality. Columns (2)–(5) of Table 2.13 present the estimates from the IV (2SLS) regressions for four different age groups when using forest cover and agricultural land ownership as IVs at the same time. The coefficient estimates on polluting fuel for cooking from the IV regressions of under-five and neonatal mortality are also positive and statistically significant at 5 percent level, ranging from 0.037 to 0.038. In other words, a

[25]The most recent development stage of theoretical models of fuel use, agricultural household models, highlights the importance of agricultural production in household fuel choice [20, 47, 61], supporting the satisfaction of relevance assumption of agricultural land ownership as a valid IV for fuel choice.

one standard-deviation increase in the use of polluting fuel is associated with increase in under-five and neonatal mortality by 3.8 and 3.7 percent, respectively. Hence, I can argue that causal effect of fuel choice on neonatal mortality drives the effect of fuel choice on under-five mortality. Compared to the 2SLS results in Table 2.12, the coefficient estimates in Table 2.13 are fairly similar. The Hansen's J-statistics suggest that the excluded IVs are exogenous, and the model is not overidentified.

2.5.2 Robustness Checks

I perform several robustness checks of my results. First, to check the robustness of my findings with respect to different estimation methods, I examine the causal effect of indoor air pollution on infant mortality using (two-step) IV probit regressions as an alternative method for IV strategy in addition to IV (2SLS) regression. I then found that IV probit regression results provide exactly the same conclusion as IV regression, verifying that the results are quite robust to different estimation approach.

Table A.1 presents the parameter estimates derived from the IV probit regression for under-five, child, post-neonatal, and neonatal mortality (with the same specification as used in Table 2.10).[26] In addition, Table A.2 presents the estimates from the IV probit regressions with specification corresponding to the IV (2SLS) regression shown in Table 2.12, using only household's agricultural land ownership as an IV for cooking fuel choice. Finally, I estimate the same

[26] I report the estimated coefficients rather than the associated average marginal effects because statistical significance of marginal effect might not be the same as that of coefficient. In other words, the inference should be about the coefficients, not partial effects, since in the latter case, one is testing a hypothesis about a function of all the coefficients, not an individual coefficient [38].

specification by adding forest cover as a second IV for polluting fuel in addition to agricultural land ownership of the household, and Table A.3 reports the estimates on the causal effect of polluting fuel use on infant mortality found from the IV probit regressions. Table A.3 shows that using dirty fuels instead of clean fuels causes under-five and neonatal mortality, and coefficient estimate on polluting fuel for cooking ranges from 0.431 to 0.589. The Wald test of exogeneity suggests that these two IVs are exogenous and there is no endogeneity after addressing the potential endogeneity using the two instruments. I also conducted non-IV analysis using OLS regression rather than probit model and found that OLS results are very similar and consistent with the estimated marginal impacts from the simple multivariate probit regressions for all four age groups (Table A.4).

Second, the government of India initiated the National Programme on Improved Chulha (NPIC) in 1984 to provide efficient cooking stoves to rural areas in an attempt to limit air pollution. NPIC became a nation-wide program in 1986 and has been implemented until 2000. In addition, a National Biomass Cookstoves Initiative (NBCI) was launched by Indian government to enhance the use of improved biomass cookstoves in 2009. In terms of coverage range, NBCI has not became a national-level program yet, and the pilot projects distributed 12,000 improved cookstoves to households in the states of Jammu and Kashmir, Uttar Pradesh, Bihar, Madhya Pradesh, Jharkhand, Chhattisgarh, Karnataka and Odisha. The program will cover all the states in the next phase.

A proper utilization of improved cook stoves is intended to provide a quick solution to short-lived pollutants by reducing smoke exposure of dirty fuels and its climate impact. A solid fuel firing household is likely to emit less pollution when using improved cooking stoves under the program. Hence, I additionally control for states where improved cookstoves program has been implemented by adding a dummy variable which indicates states where there is NBCI program. It is important to note that I did not control for states with NPIC since it already became a nationally disseminated program. Table A.5 shows the estimated average marginal impacts of cooking fuel choice on infant mortality for four-different age groups from multivariate probit regressions. Table A.6 reports results from the first and second-stage regressions of IV (2SLS) regression with a dummy variable is added to my preferred specifications in Table 2.13. The robustness check of my findings with respect to an additional control for cookstoves program states suggests that including a dummy variable indicating states where NBCI has been implemented by the government of India does not change the simple probit and IV (2SLS) regression results at all, showing that key results are robust to inclusion of this additional control.

Third, I make the key regressor a continuous variable by ranking fuel types from 1 (the cleanest fuel) to 10 (the dirtiest fuel) based on their cleanliness or energy ladder concept. I explore this as an alternative robustness check of my results. In details, I assign values to different types of fuels used for cooking as follows: 1 = electricity, 2 = LPG/natural gas, 3 = biogas, 4 = kerosene, 5 = coal/lignite, 6 = charcoal, 7 = wood, 8 = straw/shrubs/grass, 9 = agricultural crop, and 10 = animal dung. The simple OLS regression results suggest that under-five and neonatal mortality increase as fuel used for cooking becomes dirtier (Tables A.7–A.10). Note that although the coefficient estimate on cooking fuel variable is statistically significant at 1% level, the value of the coefficient is negligibly small. Table A.11 presents the IV (2SLS) regression results when a continuous variable which measures the dirtiness of cooking fuels is used as the key regressor and district-wise forest cover and a dummy variable for household's agricultural land ownership are included as IVs. The both instruments have positive and statistically significant effect on dirtiness level of cooking fuels, which is consistent with my main results in Column (1) of Table 2.13. Columns (2)–(5) of Table A.11 shows that a one standard-deviation increase in the dirtiness of cooking fuels increases the under-five and neonatal mortality by 0.6 and 0.5 percent, respectively. Since my main conclusion stays the same as before, I say that my results are robust in terms of different type of variable for polluting fuel for cooking, a continuous variable measuring the dirtiness of cooking fuels.

2.6 Conclusion

Although the world has been making progress in shifting from solid fuels to modern energy sources since 1980s, almost half of the global population still relies on solid fuels for cooking, the largest source of indoor air pollution (IAP). The IAP is a major global health threat and the leading environmental factor for death, nearly 4 million people die from illness attributable to IAP annually. India has been leading the way among major countries in IAP by extensively depending on solid fuels. By 2015, 64% of Indian population uses different types of solid fuels for cooking including wood, dung and coal, after Sub-Saharan Africa. Each year, diseases attributed to IAP kill 1.2 million people, of which 100 thousand children in India.

While previous literature has examined the effectiveness of particular government policies and programs intended to reduce the IAP (e.g., improved cooking stoves) and the effect of diseases attributable to the IAP on deaths, there is limited evidence regarding the impact of cooking fuels on child mortality. Leveraging a unique and large-scale household survey data and geospatial information of forest cover from 1992 to 2016 in India, I find that the use of solid fuels for cooking increases under-five mortality from a variety of empirical specifications. My findings extend the naïve ordinary least squares (OLS) and logistic regression estimates of the impact of IAP on child mortality.

The analysis presents two important departures from the literature. First, I utilize the nationally-representative demographic survey data in the empirical analysis instead of focusing on RCTs conducted in particular region of the country as commonly analyzed in the literature [25, 43, 78], i.e., the sample used in my empirical analysis is strictly representative of all of India. Analyses based on simple probit regressions lead to 0.3-0.6 percentage points decrease in the estimates of marginal impact of cooking fuel on infant mortality. This suggests that the literature tended to overestimate the association between IAP and infant mortality, particularly, their suggested number of under-five mortality incidences due to dirty cooking fuels use is higher than my estimates by approximately 76,000-152,000 deaths per year nationally. Since my non-IV estimation departs from literature only in terms of additional controls and larger (the most recent) sample, adding controls and utilizing larger sample decrease the coefficient estimate on fuel choice.

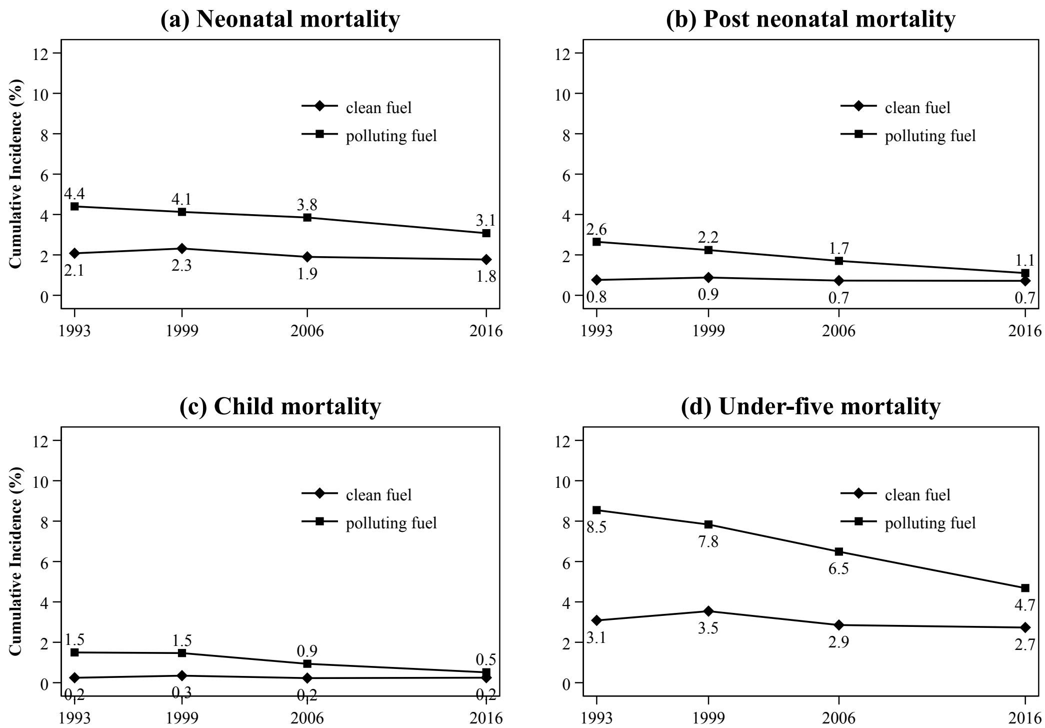
Second, this is the first empirical analysis to address the endogeneity issue due to simultaneity when quantifying the causal effect of cooking fuels on infant mortality. The forest cover and agricultural land ownership in India provide plausibly exogenous variation in cooking fuels for the causal identification.

Analysis by addressing endogeneity in the cooking fuel and infant mortality relationship shows that use of polluting fuels for cooking and resulting poor indoor air quality significantly increases the likelihood of under-five and neonatal mortality. In particular, the IV (2SLS) analysis based on forest cover and agricultural land ownership shows that a one standard-deviation increase in the use of polluting fuel increases the neonatal and under-five mortality by 3.7 and 3.8 percent, respectively, which are higher estimates compared to simple OLS regression results. Moreover, I found no causal impact of fuel choice on post-neonatal and child mortality. Therefore, cooking fuel choice affects under-five mortality mainly through neonatal mortality since mortality for all three preceding age groups (neonatal, post-neonatal, and child) are completely inclusive in under-five mortality.

I conclude with some caveats and directions for future research. First, my analysis is based on an indirect indicator of IAP, i.e., type of cooking fuels, to estimate the effect of IAP on under-five mortality due to the lack of data availability. Using direct measures of IAP (CO and PM emissions in homes) recorded by 24-hour carbon monoxide readings might provide more accurate estimation. Although cooking is the main source of IAP, it is not the only source of CO emission inside the house that poses risks to children's health. The WHO guidelines for household fuel combustion [86] classify kerosene as a polluting fuel and discourage its use as a household fuel. Nevertheless, kerosene is still used for lighting by around one billion people who lack access to electricity. Kerosene lamps are often the only means of lighting houses at night, allowing children and adolescents to study in areas without electricity. Use of kerosene not only pollutes the air inside the house but also increases the risks for fires, burns and CO poisoning. Therefore, since the use of polluting fuel for cooking is the largest source of IAP, I might have underestimated the effect of IAP on infant mortality due to the absence of direct measure of IAP and indirect measures for other sources of household air pollution.

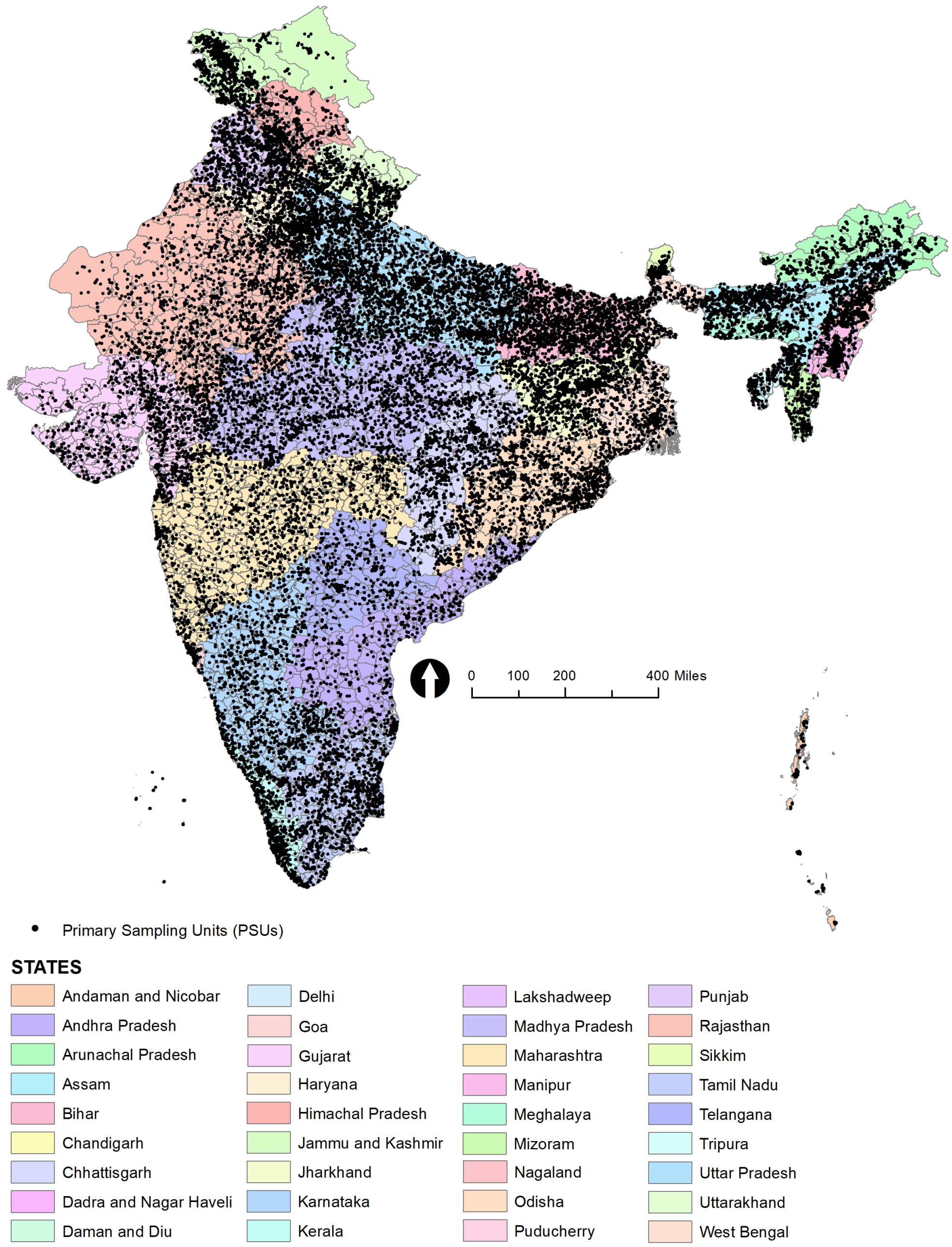
Second, I focus on the causal impact of IAP on infant mortality. It is well understood that IAP affects not only infant mortality but also other socio-economic and health outcomes. Hence, future research could empirically examine the impact of cooking fuels on productivity of children and adults, school attendance, labor market participation, all of which could have important implications on the broader economy and contribute to the economic literature of indoor air quality or fuel choice.

Figure 2.1: Mortality Trend in All Age-Groups of Children Under-Five associated with IAP in India



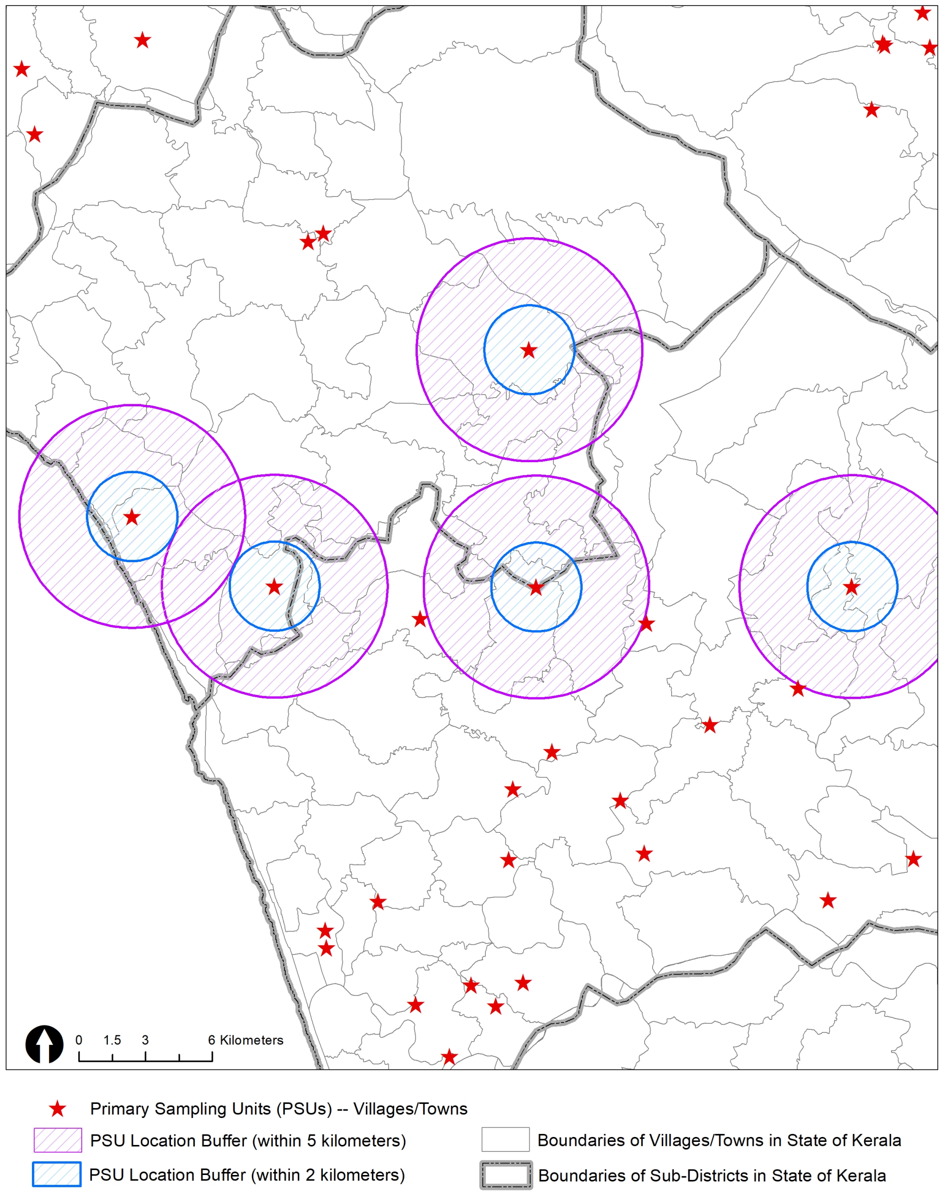
Note: The figure presents the trend of infant mortality rates (cumulative incidence proportion) in India for three different age-groups (neonatal, post-neonatal, and child) and overall under-five over the period 1992–2016. The NFHS datasets for the period 1992–93, 1998–99, 2005–06, and 2015–16 was used to calculate the mortality rates. To adjust for the cluster sampling survey design, “svy” command was used for calculating weighted cumulative mortality incidences. The trend of mortality incidence proportion was analyzed in relation to the types of fuels used for cooking (clean or polluting fuels).

Figure 2.2: Distribution of PSUs (Villages/City Blocks) in India's NFHS-4 (2015–16)



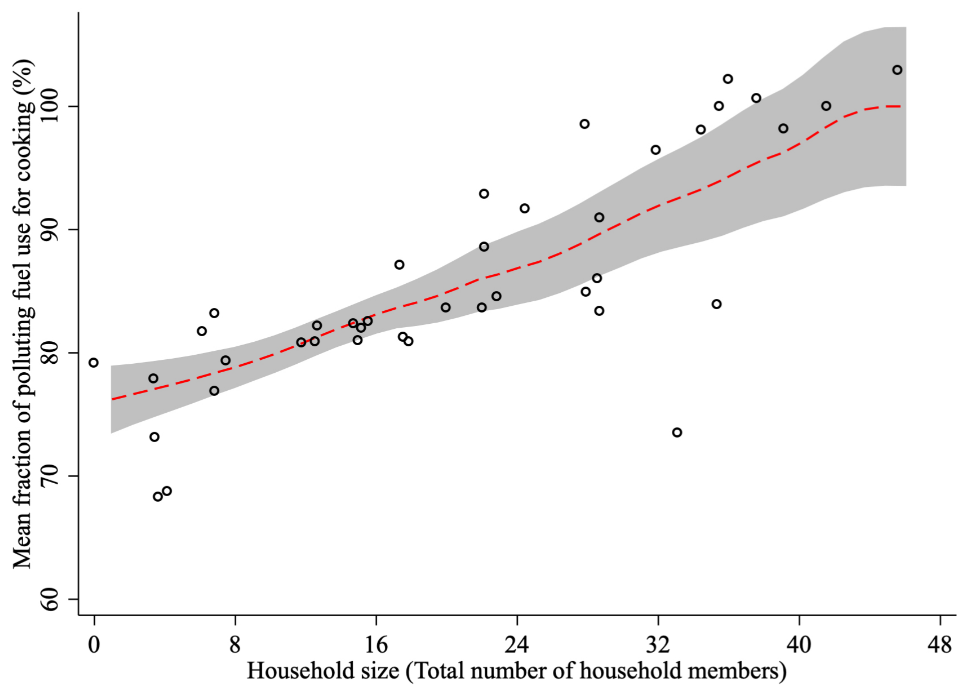
Note: The figure plots the primary sampling units (PSUs)–villages in rural areas and city blocks in urban areas–in India's NFHS-4 (2015–16). The GPS information of PSUs is only available for the fourth-round of the survey as it was the first time collecting Geo-codes in NFHS-4 and any identification of PSUs including the GPS data is not available in any of the earlier rounds. A total of 28,526 points are plotted in this map or the survey is collected from households residing in 28,526 PSUs (clusters), which is the smallest geographical survey statistical unit for NFHS surveys.

Figure 2.3: Displacement of PSUs (Villages/City Blocks) in India's NFHS-4 (2015–16)



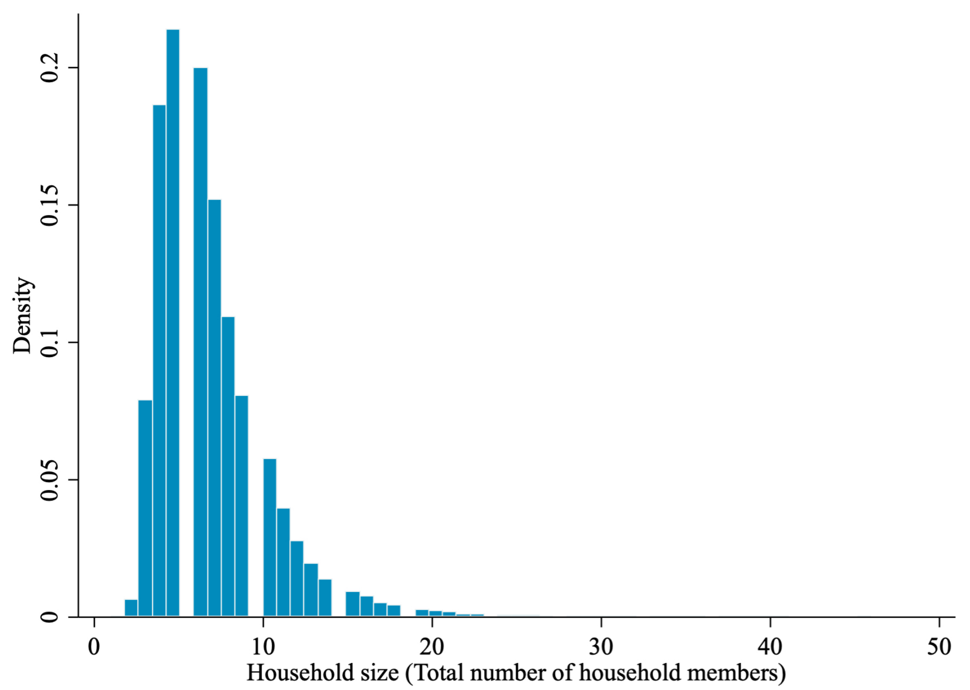
Note: The figure shows how the PSU points are displaced in NFHS-4 (2015–16) survey based on few PSU points in Kerala district. In order to ensure that respondent confidentiality is maintained, the GPS (latitude/longitude positions) of respondent locations are randomly displaced according to the "random direction, random distance" method. The displacement is randomly carried out so that (i) urban clusters are displaced up to 2 kilometers, (ii) rural clusters are displaced up to 5 kilometers, with 1% of the rural clusters displaced up to 10 kilometers. According to the description of the DHS GPS data provided by the DHS Program, the displacement is restricted so that the points stay within the same country, state, and district areas as the undisplaced cluster. The buffer analysis on few PSU points in Kerala district as an example suggests that identifications of villages/towns and sub-districts are questionable because 2-5-kilometer buffers intersect with boundaries of villages/towns and sub-districts.

Figure 2.4: Relationship between Household Size and Fuel Choice



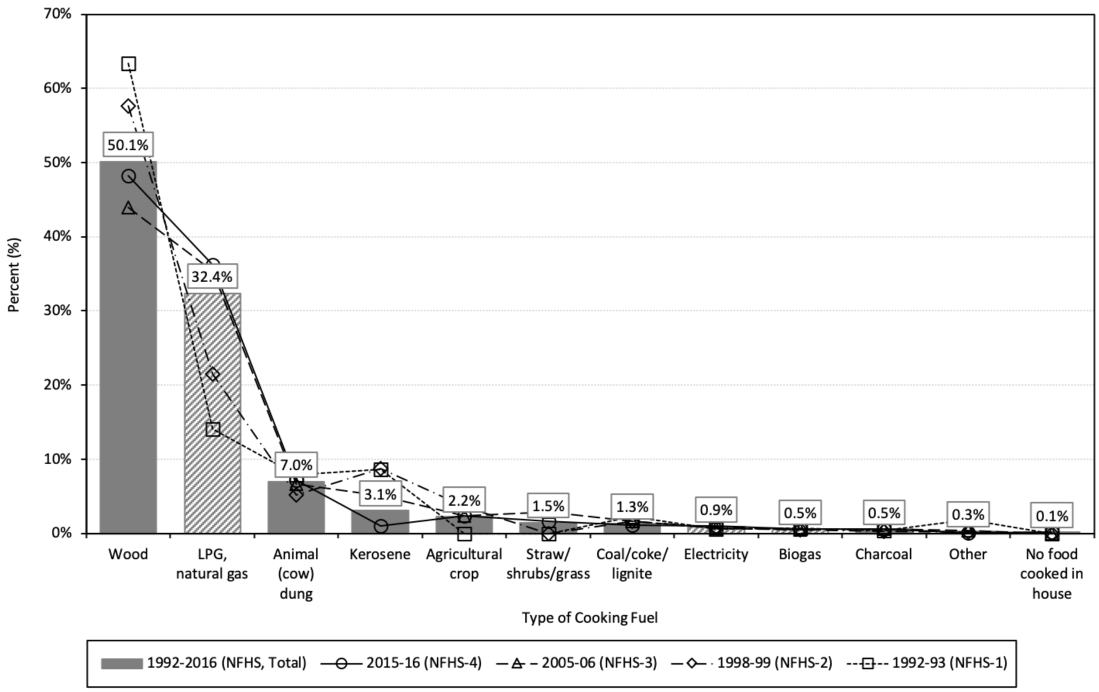
Note: The figure plots association between tendency of polluting fuel use for cooking and household size (or total number of household members) based on NFHS data, after removing few outliers, and it suggests that there is a strong and positive relationship between these two variables. Thus, I argue that household size is clearly one of the important factors that affect households' fuel choice. The scatter plot also reveals that a threshold household size which causes shift (or kink) in type of cooking fuel could be around 8 people.

Figure 2.5: Histogram of Household Size in the NFHS Data



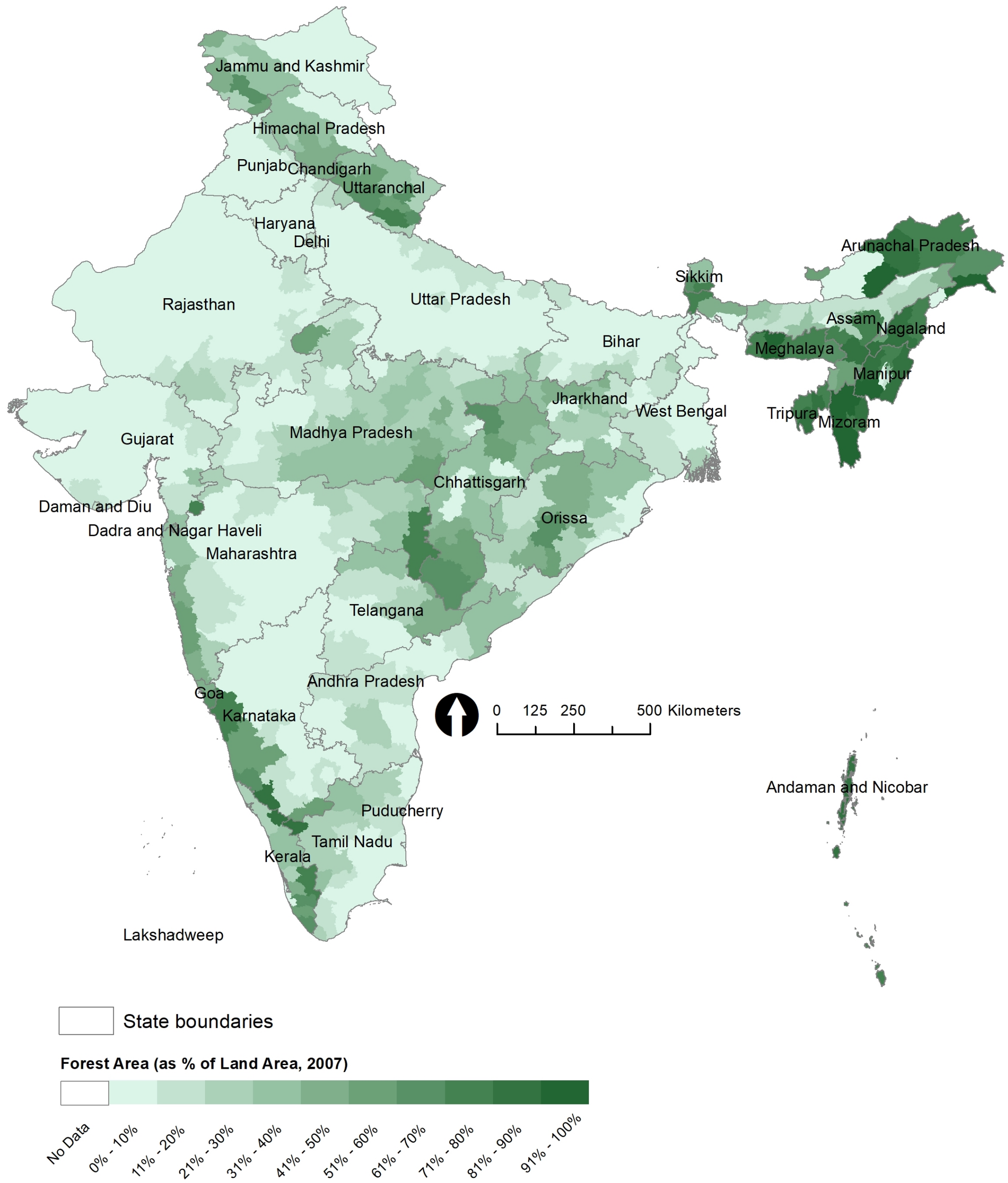
Note: The figure depicts a histogram of how many households in the sample by household size. While Figure 2.4 suggests that all households with members above 32 people use only polluting fuels, this figure shows that those type of households account for very small fraction in total households covered in the NFHS.

Figure 2.6: Share of Households in the NFHS relying on Different Types of Fuels for Cooking



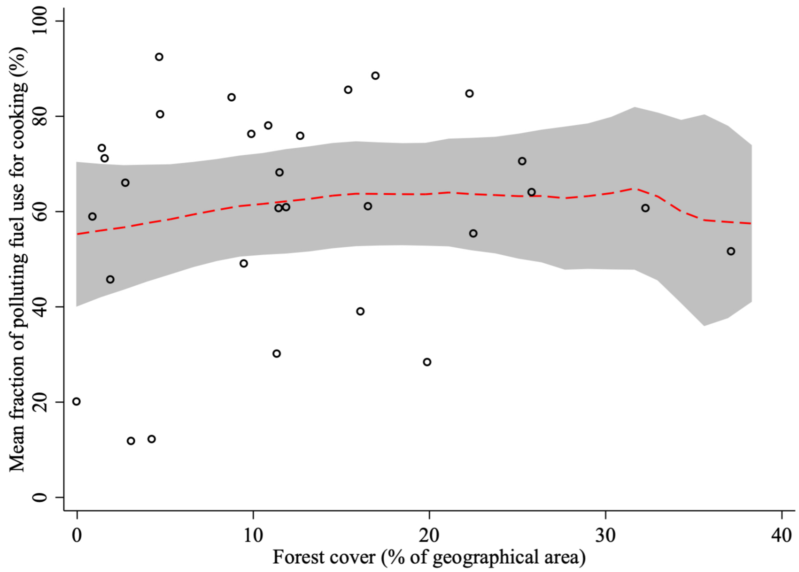
Note: The figure shows the share of households covered in four rounds of National Family Health Survey (NFHS) using different types of fuels for cooking in India over the period 1992–2016. The line charts depict the share of households using each type of cooking fuel for individual rounds of survey, while the bar chart illustrates the share for all four rounds of survey between 1992 and 2016 (the bars for clean fuels are filled with pattern, whereas the bars for polluting fuels are in solid fill). Wood is the leading fuel used for cooking in India, accounting for 50.1% of the sampling households in the NFHS over the period 1992–2016. The second dominant cooking fuel is a liquid petroleum gas (LPG) and/or natural gas with the share of 32.4%. The other clean fuels account for only 1.4% (electricity and biogas account for about 0.9% and 0.4%, respectively). Overall, based on my classification of cooking fuels, I can see that one-third of the Indian households have been consuming clean fuels for their cooking, while the majority or the remaining two-third of the households have been relying on polluting fuels for cooking over the past 25 years.

Figure 2.7: India's District-Wise Forest Cover

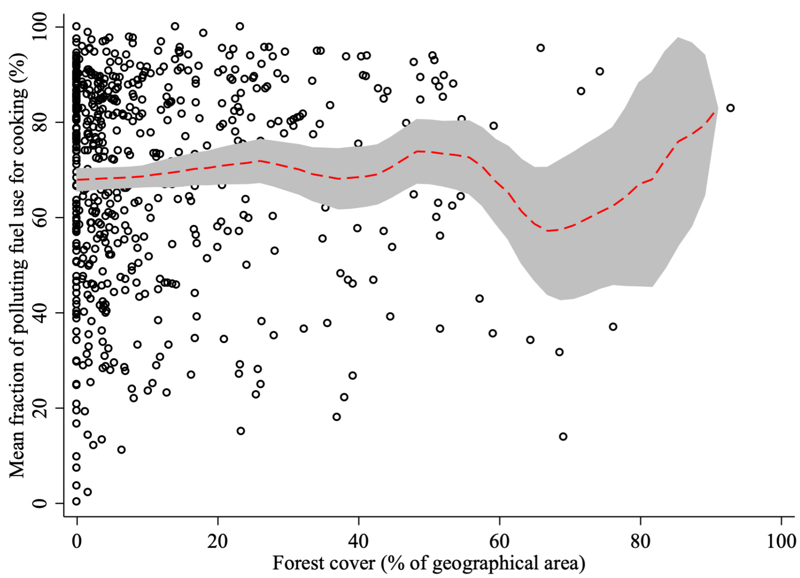


Note: The figure depicts the 2007 district-wise forest cover (measured by percentage of geographical area covered by forests) in India. The forest cover includes all types of forests (different canopy density classes) including very dense, moderately dense, and open forests.

Figure 2.8: Relationship between Cooking Fuel Choice and Forest Cover



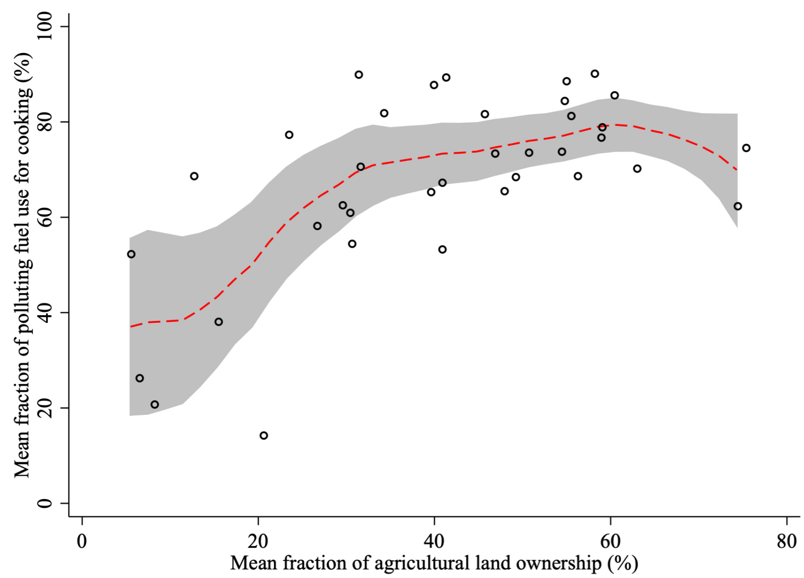
(a) State-level relationship



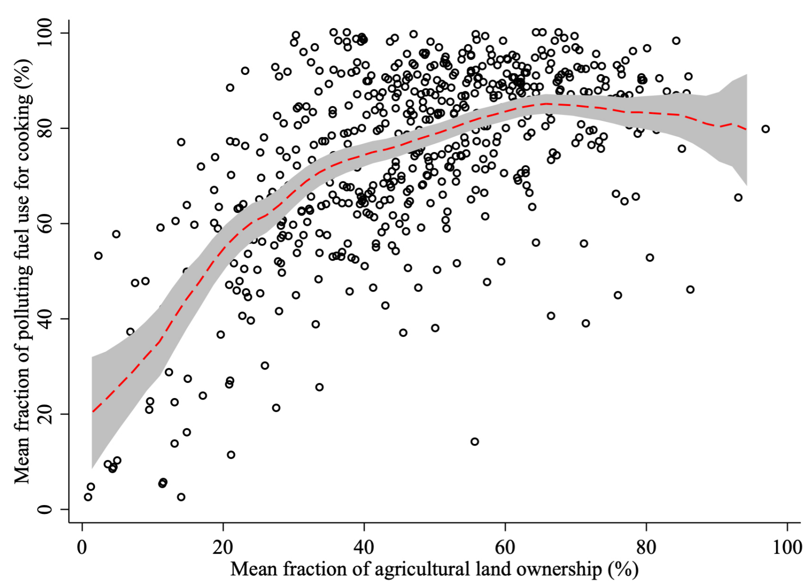
(b) District-level relationship

Note: The figure plots the observations of the variables pairwise at (a) state-level and (b) district-level: forest cover (% of geographical area) versus mean fraction of polluting fuel use for cooking (%). The scatter plot was depicted for 30 states and 627 districts of India (after excluding few outliers). The correlation coefficient between mean fraction of polluting fuel use for cooking and forest cover at the state level is 0.11, while bivariate correlation between the two indicators at district-level is 0.02.

Figure 2.9: Relationship between Cooking Fuel Choice and Agricultural Land Ownership



(a) State-level relationship



(b) District-level relationship

Note: The figure plots the observations of the variables pairwise at (a) state-level and (b) district-level: mean fraction of agricultural land ownership (%) versus mean fraction of polluting fuel use for cooking (%). The scatter plot was depicted for 36 states and 603 districts of India. The correlation coefficient between mean fractions of polluting fuel use for cooking and agricultural land ownership at the state level is 0.63, while bivariate correlation between the two indicators at district-level is 0.59.

Table 2.1: Summary Statistics

Variables	Mean	S.D.	Min	Max	N
Infant mortality					
Under-five	0.053	0.224	0.000	1.000	369,416
Child	0.007	0.085	0.000	1.000	369,416
Post-neonatal	0.015	0.120	0.000	1.000	369,416
Neonatal	0.031	0.173	0.000	1.000	369,416
Type of cooking fuel					
Clean	0.231	0.421	0.000	1.000	369,416
Polluting	0.769	0.421	0.000	1.000	369,416
Place of residence					
Urban	0.244	0.430	0.000	1.000	369,416
Rural	0.756	0.430	0.000	1.000	369,416
Household income (wealth index)					
High	0.150	0.357	0.000	1.000	369,416
Middle	0.385	0.487	0.000	1.000	369,416
Low	0.465	0.499	0.000	1.000	369,416
Mother's age					
40-49	0.027	0.162	0.000	1.000	369,416
<20	0.041	0.199	0.000	1.000	369,416
20-29	0.679	0.467	0.000	1.000	369,416
30-39	0.253	0.435	0.000	1.000	369,416
Mother's education					
Secondary/Higher	0.458	0.498	0.000	1.000	369,219
Primary	0.151	0.358	0.000	1.000	369,219
No education	0.392	0.488	0.000	1.000	369,219
Gender of child					
Female	0.481	0.500	0.000	1.000	369,416
Male	0.519	0.500	0.000	1.000	369,416
Breastfeeding status					
Ever breastfed	0.654	0.476	0.000	1.000	369,416
Never breastfed	0.346	0.476	0.000	1.000	369,416
Place where food is cooked					
In same room as they live in	0.369	0.483	0.000	1.000	251,983
In separate kitchen inside the house	0.446	0.497	0.000	1.000	251,983
In a separate building	0.106	0.308	0.000	1.000	251,870
Outdoors	0.078	0.269	0.000	1.000	251,870
Type of house					
Pucca	0.376	0.484	0.000	1.000	358,410
Semi-pucca	0.437	0.496	0.000	1.000	358,410
Kachha	0.187	0.390	0.000	1.000	358,410
Number of household members	6.864	3.253	1.000	46.000	369,416

Note: The table summarizes the household and individual characteristics of respondents from three rounds of NFHS (1992–93, 1998–99, and 2015–16) used in the regression analysis. The unit of observation is the child. Neonatal = first 28 days of life (0–28 days), Post-neonatal = period between approximately the first month after birth and end of the first year of life (1-11 months), and Child = period between exact ages of one and five (12–59 months).

Table 2.2: Summary Statistics of Infant Mortality & Fuel Choice (by State)

States	Infant mortality (fraction)							
	Under-Five		Child		Post-Neonatal		Neonatal	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Andaman and Nicobar Islands	0.008	0.088	0.000	0.000	0.003	0.056	0.005	0.068
Andhra Pradesh	0.055	0.228	0.006	0.076	0.016	0.125	0.033	0.179
Arunachal Pradesh	0.036	0.186	0.010	0.101	0.012	0.110	0.014	0.115
Assam	0.058	0.234	0.008	0.092	0.016	0.127	0.033	0.179
Bihar	0.059	0.236	0.008	0.088	0.014	0.118	0.037	0.190
Chandigarh	0.036	0.188	0.000	0.000	0.005	0.072	0.031	0.174
Chhattisgarh	0.058	0.235	0.006	0.08	0.012	0.110	0.040	0.196
Dadra and Nagar Haveli	0.028	0.166	0.003	0.056	0.016	0.124	0.009	0.097
Daman and Diu	0.030	0.172	0.003	0.050	0.008	0.087	0.020	0.141
Delhi	0.050	0.219	0.006	0.075	0.019	0.137	0.026	0.158
Goa	0.027	0.163	0.002	0.046	0.006	0.08	0.019	0.136
Gujarat	0.050	0.218	0.008	0.089	0.013	0.112	0.029	0.168
Haryana	0.046	0.209	0.007	0.082	0.014	0.119	0.025	0.155
Himachal Pradesh	0.038	0.192	0.004	0.062	0.011	0.106	0.023	0.150
Jammu and Kashmir	0.041	0.198	0.005	0.068	0.011	0.105	0.025	0.156
Jharkhand	0.045	0.208	0.006	0.077	0.010	0.100	0.029	0.168
Karnataka	0.042	0.200	0.005	0.070	0.010	0.102	0.027	0.161
Kerala	0.015	0.121	0.002	0.040	0.004	0.063	0.009	0.096
Lakshadweep	0.014	0.116	0.000	0.000	0.003	0.058	0.010	0.101
Madhya Pradesh	0.066	0.249	0.012	0.108	0.017	0.129	0.038	0.190
Maharashtra	0.033	0.178	0.004	0.064	0.007	0.085	0.021	0.144
Manipur	0.028	0.166	0.004	0.064	0.007	0.085	0.017	0.129
Meghalaya	0.050	0.218	0.008	0.086	0.018	0.132	0.025	0.156
Mizoram	0.044	0.206	0.005	0.074	0.027	0.163	0.012	0.108
Nagaland	0.034	0.180	0.005	0.073	0.013	0.115	0.015	0.122
Odisha	0.060	0.237	0.007	0.083	0.019	0.136	0.034	0.181
Puducherry	0.015	0.122	0.003	0.053	0.004	0.061	0.009	0.092
Punjab	0.042	0.200	0.005	0.069	0.012	0.110	0.025	0.155
Rajasthan	0.059	0.235	0.009	0.096	0.017	0.129	0.033	0.177
Sikkim	0.037	0.188	0.005	0.068	0.012	0.107	0.020	0.141
Tamil Nadu	0.035	0.184	0.005	0.074	0.009	0.094	0.021	0.143
Telangana	0.042	0.200	0.004	0.062	0.010	0.098	0.028	0.165
Tripura	0.045	0.208	0.004	0.063	0.018	0.133	0.023	0.151
Uttar Pradesh	0.074	0.262	0.010	0.099	0.021	0.142	0.044	0.204
Uttarakhand	0.045	0.207	0.004	0.067	0.013	0.112	0.028	0.165
West Bengal	0.047	0.211	0.006	0.080	0.011	0.105	0.029	0.167
Total	0.053	0.224	0.007	0.085	0.015	0.120	0.031	0.173

Table 2.2: (Continued)

States	Type of cooking fuel (fraction)			N
	Mean		S.D.	
	Clean	Polluting		
Andaman and Nicobar Islands	0.478	0.522	0.500	638
Andhra Pradesh	0.333	0.667	0.471	5,515
Arunachal Pradesh	0.306	0.694	0.461	6,517
Assam	0.144	0.856	0.351	14,393
Bihar	0.091	0.909	0.288	33,093
Chandigarh	0.839	0.161	0.369	192
Chhattisgarh	0.153	0.847	0.360	10,695
Dadra and Nagar Haveli	0.453	0.547	0.499	320
Daman and Diu	0.716	0.284	0.451	395
Delhi	0.653	0.347	0.476	5,264
Goa	0.410	0.590	0.492	2,339
Gujarat	0.265	0.735	0.441	12,077
Haryana	0.316	0.684	0.465	11,795
Himachal Pradesh	0.259	0.741	0.438	6,180
Jammu and Kashmir	0.342	0.658	0.474	11,059
Jharkhand	0.104	0.896	0.305	13,427
Karnataka	0.289	0.711	0.453	12,462
Kerala	0.255	0.745	0.436	5,545
Lakshadweep	0.286	0.714	0.453	294
Madhya Pradesh	0.181	0.819	0.385	32,007
Maharashtra	0.333	0.667	0.471	14,581
Manipur	0.289	0.711	0.454	7,341
Meghalaya	0.104	0.896	0.305	6,261
Mizoram	0.469	0.531	0.499	6,257
Nagaland	0.170	0.830	0.375	5,672
Odisha	0.113	0.887	0.317	16,192
Puducherry	0.818	0.182	0.386	1,057
Punjab	0.407	0.593	0.491	8,426
Rajasthan	0.180	0.820	0.384	25,435
Sikkim	0.310	0.690	0.463	1,717
Tamil Nadu	0.464	0.536	0.499	11,849
Telangana	0.393	0.607	0.488	4,142
Tripura	0.142	0.858	0.349	2,542
Uttar Pradesh	0.195	0.805	0.396	56,090
Uttarakhand	0.312	0.688	0.463	7,601
West Bengal	0.122	0.878	0.327	10,046
Total	0.231	0.769	0.421	369,416

Note: The table summarizes the infant mortality of four different age-groups (outcome variables) and the type of cooking fuel (key explanatory variable) by state recorded in three rounds of NFHS (1992–93, 1998–99, and 2015–16) used in the regression analysis. All 35 regions of India (29 states and six union territories) are considered. Infant mortality and fuel choices significantly vary across regions throughout the country.

Table 2.3: Summary Statistics of Infant Mortality & Fuel Choice (by Age of the Household Head)

	Infant mortality (fraction)								Type of cooking fuel (fraction)			N
	Under-Five		Child		Post-Neonatal		Neonatal		Mean		S.D.	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Clean	Polluting		
Age 10-19	0.100	0.301	0.012	0.107	0.041	0.197	0.048	0.215	0.066	0.934	0.248	518
Age 20-29	0.058	0.235	0.007	0.082	0.016	0.125	0.036	0.186	0.186	0.814	0.389	62,807
Age 30-39	0.052	0.222	0.009	0.095	0.015	0.123	0.028	0.164	0.245	0.755	0.430	111,513
Age 40-49	0.061	0.239	0.010	0.099	0.017	0.127	0.034	0.182	0.196	0.804	0.397	53,663
Age 50-59	0.049	0.217	0.005	0.072	0.013	0.114	0.031	0.173	0.245	0.755	0.430	60,828
Age 60-69	0.046	0.209	0.005	0.069	0.012	0.109	0.029	0.168	0.263	0.737	0.440	56,211
Age 70-79	0.048	0.213	0.006	0.077	0.013	0.111	0.029	0.168	0.252	0.748	0.434	18,920
Age 80-89	0.052	0.221	0.006	0.076	0.016	0.126	0.030	0.170	0.229	0.771	0.420	4,254
Age $\geq$ 90	0.070	0.256	0.008	0.088	0.020	0.141	0.042	0.201	0.191	0.809	0.393	640
Total	0.053	0.224	0.007	0.085	0.015	0.120	0.031	0.173	0.231	0.769	0.421	369,354

Note: The table summarizes the infant mortality of four different age-groups (outcome variables) and the type of cooking fuel (key explanatory variable) for different age groups of the household head recorded in three rounds of NFHS (1992–93, 1998–99, and 2015–16) used in the regression analysis. Infant mortality and fuel choices are generally the same across different age groups of the household head, suggesting that age of the household does not affect the outcome variables and the key regressor.

Table 2.4: Summary Statistics of Infant Mortality & Fuel Choice (by Gender of the Household Head)

	Infant mortality (fraction)								Type of cooking fuel (fraction)			N
	Under-Five		Child		Post-Neonatal		Neonatal		Mean		S.D.	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Clean	Polluting		
Male	0.053	0.225	0.008	0.086	0.015	0.121	0.031	0.174	0.229	0.771	0.420	330,884
Female	0.048	0.214	0.006	0.077	0.014	0.117	0.028	0.166	0.247	0.753	0.431	38,530
Total	0.053	0.224	0.007	0.085	0.015	0.120	0.031	0.173	0.231	0.769	0.421	369,414

Note: The table summarizes the infant mortality of four different age-groups (outcome variables) and the type of cooking fuel (key explanatory variable) by gender of the household head recorded in three rounds of NFHS (1992–93, 1998–99, and 2015–16) used in the regression analysis. The summary statistic suggests that infant mortality and fuel choices do not depend on gender of the household head.

Table 2.5: Probit: The Marginal Impact of Cooking Fuel Choice on Under-Five Mortality

	Dependent variable: Under-five mortality				
	(1)	(2)	(3)	(4)	(5)
Polluting fuel for cooking	0.033***	0.010***	0.010***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Place of residence: Rural		0.003**	0.003**	0.003***	0.002*
		(0.001)	(0.001)	(0.001)	(0.001)
Household income: Middle		0.008***	0.008***	0.012***	0.011***
		(0.002)	(0.002)	(0.002)	(0.002)
Household income: Low		0.013***	0.013***	0.016***	0.014***
		(0.002)	(0.002)	(0.002)	(0.002)
Number of household members		-0.003***	-0.003***	-0.004***	-0.004***
		(0.000)	(0.000)	(0.000)	(0.000)
Mother's age: <20		0.016***	0.017***	0.016***	0.016***
		(0.003)	(0.003)	(0.003)	(0.003)
Mother's age: 20-29		-0.006**	-0.006**	-0.008***	-0.007***
		(0.002)	(0.002)	(0.002)	(0.002)
Mother's age: 30-39		-0.012***	-0.011***	-0.013***	-0.013***
		(0.002)	(0.002)	(0.002)	(0.002)
Mother's education: Primary		0.011***	0.011***	0.009***	0.008***
		(0.001)	(0.001)	(0.001)	(0.001)
Mother's education: No education		0.014***	0.014***	0.010***	0.009***
		(0.001)	(0.001)	(0.001)	(0.001)
Gender of child: Male		0.004***	0.004***	0.004***	0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
Never breastfed		0.049***	0.049***	0.050***	0.050***
		(0.001)	(0.001)	(0.001)	(0.001)
Food cooked: In separate kitchen inside		-0.007***	-0.007***	-0.003**	-0.003***
		(0.001)	(0.001)	(0.001)	(0.001)
Food cooked: In a separate building		-0.009***	-0.009***	-0.003*	-0.002
		(0.001)	(0.001)	(0.002)	(0.002)
Food cooked: Outdoors		-0.005***	-0.005***	-0.000	-0.000
		(0.002)	(0.002)	(0.002)	(0.002)
House type: Semi-pucca		0.006***	0.007***	0.004***	0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
House type: Kachha		0.005***	0.005***	0.005***	0.005***
		(0.002)	(0.002)	(0.002)	(0.002)
Year FE	N	N	Y	Y	Y
State-by-Year FE	N	N	N	Y	N
District-by-Year FE	N	N	N	N	Y
N	369,416	242,971	242,971	242,971	241,921
Probit log-likelihood	-75,772	-39,646	-39,635	-39,207	-38,701

Note: Each column reports average marginal effects for a multivariate probit regression where the dependent variable is under-five mortality and the key explanatory variable is indoor air pollution (polluting fuel for cooking). The year fixed effects include dummies for years of interview. The state and district fixed effects include dummies for 36 states and 603 districts, respectively, and the interactions term include the interactions of state and district dummies with year-of-interview fixed effects (State-by-Year FE and District-by-Year FE). The number of observations slightly dropped in Column (5) due to some missing districts. The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 2.6: Probit: The Marginal Impact of Cooking Fuel Choice on Child Mortality

	Dependent variable: Child mortality				
	(1)	(2)	(3)	(4)	(5)
Polluting fuel for cooking	0.009***	0.001***	0.001***	0.001***	0.002***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Place of residence: Rural		0.001	0.001	0.001	0.001*
		(0.000)	(0.000)	(0.000)	(0.001)
Household income: Middle		0.001	0.001	0.001*	0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Household income: Low		0.002**	0.002**	0.002**	0.002**
		(0.001)	(0.001)	(0.001)	(0.001)
Number of household members		-0.001***	-0.001***	-0.001***	-0.001***
		(0.000)	(0.000)	(0.000)	(0.000)
Mother's age: <20		-0.004***	-0.004***	-0.004***	-0.005***
		(0.001)	(0.001)	(0.001)	(0.001)
Mother's age: 20-29		-0.003***	-0.003***	-0.003***	-0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
Mother's age: 30-39		-0.002***	-0.002***	-0.002***	-0.002***
		(0.001)	(0.001)	(0.001)	(0.001)
Mother's education: Primary		0.002***	0.002***	0.002***	0.003***
		(0.000)	(0.000)	(0.000)	(0.001)
Mother's education: No education		0.004***	0.004***	0.003***	0.004***
		(0.000)	(0.000)	(0.000)	(0.000)
Gender of child: Male		-0.001***	-0.001***	-0.001***	-0.001***
		(0.000)	(0.000)	(0.000)	(0.000)
Never breastfed		0.004***	0.004***	0.004***	0.005***
		(0.000)	(0.000)	(0.000)	(0.000)
Food cooked: In separate kitchen inside		-0.001***	-0.001***	-0.001*	-0.001**
		(0.000)	(0.000)	(0.000)	(0.000)
Food cooked: In a separate building		-0.001**	-0.001**	-0.001	-0.001
		(0.000)	(0.000)	(0.001)	(0.001)
Food cooked: Outdoors		-0.000	-0.000	-0.000	-0.000
		(0.001)	(0.001)	(0.001)	(0.001)
House type: Semi-pucca		0.001***	0.001***	0.001**	0.001
		(0.000)	(0.000)	(0.000)	(0.001)
House type: Kachha		0.001*	0.001*	0.001	0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Year FE	N	N	Y	Y	Y
State-by-Year FE	N	N	N	Y	N
District-by-Year FE	N	N	N	N	Y
N	369,416	242,971	242,971	241,259	192,456
Probit log-likelihood	-15,812	-6,620	-6,620	-6,575	-6,225

Note: Each column reports average marginal effects for a multivariate probit regression where the dependent variable is child mortality and the key explanatory variable is indoor air pollution (polluting fuel for cooking). The year fixed effects include dummies for years of interview. The state and district fixed effects include dummies for 36 states and 603 districts, respectively, and the interactions term include the interactions of state and district dummies with year-of-interview fixed effects (State-by-Year FE and District-by-Year FE). The number of observations slightly dropped in Column (5) due to some missing districts. The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 2.7: Probit: The Marginal Impact of Cooking Fuel Choice on Post-Neonatal Mortality

	Dependent variable: Post-neonatal mortality				
	(1)	(2)	(3)	(4)	(5)
Polluting fuel for cooking	0.009***	0.001*	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Place of residence: Rural		0.001	0.001	0.001**	0.001*
		(0.001)	(0.001)	(0.001)	(0.001)
Household income: Middle		0.004***	0.004***	0.005***	0.005***
		(0.001)	(0.001)	(0.001)	(0.001)
Household income: Low		0.005***	0.005***	0.007***	0.007***
		(0.001)	(0.001)	(0.001)	(0.001)
Number of household members		-0.001***	-0.001***	-0.001***	-0.001***
		(0.000)	(0.000)	(0.000)	(0.000)
Mother's age: <20		0.002	0.002	0.003**	0.003**
		(0.002)	(0.002)	(0.002)	(0.002)
Mother's age: 20-29		-0.004***	-0.003***	-0.003***	-0.003***
		(0.001)	(0.001)	(0.001)	(0.001)
Mother's age: 30-39		-0.004***	-0.004***	-0.003***	-0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
Mother's education: Primary		0.003***	0.003***	0.002***	0.002***
		(0.001)	(0.001)	(0.001)	(0.001)
Mother's education: No education		0.004***	0.004***	0.004***	0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
Gender of child: Male		-0.001	-0.001	-0.001	-0.001
		(0.000)	(0.000)	(0.000)	(0.000)
Never breastfed		0.013***	0.013***	0.013***	0.015***
		(0.001)	(0.001)	(0.001)	(0.001)
Food cooked: In separate kitchen inside		-0.002***	-0.002***	0.000	0.000
		(0.001)	(0.001)	(0.001)	(0.001)
Food cooked: In a separate building		-0.002***	-0.002**	-0.000	-0.000
		(0.001)	(0.001)	(0.001)	(0.001)
Food cooked: Outdoors		-0.002**	-0.002*	0.000	-0.000
		(0.001)	(0.001)	(0.001)	(0.001)
House type: Semi-pucca		0.001**	0.001**	0.000	0.000
		(0.001)	(0.001)	(0.001)	(0.001)
House type: Kachha		0.001	0.001	-0.000	-0.000
		(0.001)	(0.001)	(0.001)	(0.001)
Year FE	N	N	Y	Y	Y
State-by-Year FE	N	N	N	Y	N
District-by-Year FE	N	N	N	N	Y
N	369,416	242,971	242,971	242,565	226,142
Probit log-likelihood	-28,052	-13,549	-13,531	-13,373	-12,961

Note: Each column reports average marginal effects for a multivariate probit regression where the dependent variable is post-neonatal mortality and the key explanatory variable is indoor air pollution (polluting fuel for cooking). The year fixed effects include dummies for years of interview. The state and district fixed effects include dummies for 36 states and 603 districts, respectively, and the interactions term include the interactions of state and district dummies with year-of-interview fixed effects (State-by-Year FE and District-by-Year FE). The number of observations slightly dropped in Column (5) due to some missing districts. The unit of observation is child. The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 2.8: Probit: The Marginal Impact of Cooking Fuel Choice on Neonatal Mortality

	Dependent variable: Neonatal mortality				
	(1)	(2)	(3)	(4)	(5)
Polluting fuel for cooking	0.017***	0.008***	0.008***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Place of residence: Rural		0.001	0.001	0.001	0.000
		(0.001)	(0.001)	(0.001)	(0.001)
Household income: Middle		0.003**	0.004***	0.006***	0.006***
		(0.001)	(0.001)	(0.001)	(0.001)
Household income: Low		0.006***	0.006***	0.007***	0.006***
		(0.002)	(0.002)	(0.002)	(0.002)
Number of household members		-0.002***	-0.002***	-0.002***	-0.002***
		(0.000)	(0.000)	(0.000)	(0.000)
Mother's age: <20		0.018***	0.018***	0.017***	0.018***
		(0.003)	(0.003)	(0.003)	(0.003)
Mother's age: 20-29		0.002	0.002	-0.000	0.000
		(0.002)	(0.002)	(0.002)	(0.002)
Mother's age: 30-39		-0.005***	-0.005***	-0.007***	-0.007***
		(0.002)	(0.002)	(0.002)	(0.002)
Mother's education: Primary		0.006***	0.006***	0.005***	0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
Mother's education: No education		0.006***	0.006***	0.003***	0.002**
		(0.001)	(0.001)	(0.001)	(0.001)
Gender of child: Male		0.006***	0.006***	0.005***	0.005***
		(0.001)	(0.001)	(0.001)	(0.001)
Never breastfed		0.031***	0.032***	0.032***	0.033***
		(0.001)	(0.001)	(0.001)	(0.001)
Food cooked: In separate kitchen inside		-0.005***	-0.005***	-0.002**	-0.002**
		(0.001)	(0.001)	(0.001)	(0.001)
Food cooked: In a separate building		-0.006***	-0.006***	-0.002*	-0.002
		(0.001)	(0.001)	(0.001)	(0.001)
Food cooked: Outdoors		-0.002*	-0.002*	-0.000	-0.000
		(0.001)	(0.001)	(0.001)	(0.001)
House type: Semi-pucca		0.004***	0.004***	0.003***	0.003***
		(0.001)	(0.001)	(0.001)	(0.001)
House type: Kachha		0.004**	0.004**	0.005***	0.005***
		(0.001)	(0.001)	(0.001)	(0.002)
Year FE	N	N	Y	Y	Y
State-by-Year FE	N	N	N	Y	N
District-by-Year FE	N	N	N	N	Y
N	369,416	242,971	242,971	242,971	238,640
Probit log-likelihood	-50,600	-28,415	-28,414	-28,008	-27,562

Note: Each column reports average marginal effects for a multivariate probit regression where the dependent variable is neonatal mortality and the key explanatory variable is indoor air pollution (polluting fuel for cooking). The year fixed effects include dummies for years of interview. The state and district fixed effects include dummies for 36 states and 603 districts, respectively, and the interactions term include the interactions of state and district dummies with year-of-interview fixed effects (State-by-Year FE and District-by-Year FE). The number of observations slightly dropped in Column (5) due to some missing districts. The unit of observation is child. The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 2.9: The Effect of Polluting Fuel for Cooking on Infant Mortality (Comparison of Results from Simple Nonlinear Models)

	NFHS-1-3 (1992–2006)		NFHS-4	NFHS-1-4	This
	Naz et al. (2016)	Replication	(2015–16)	(1992–2016)	paper
	Dependent variable: Under-five mortality				
Odds Ratio	1.30***	1.27***	1.32***	1.26***	1.25***
		(0.065)	(0.114)	(0.054)	(0.045)
Marginal Effect		0.014***	0.011***	0.013***	0.008***
		(0.003)	(0.003)	(0.003)	(0.001)
N	138,063	150,845	34,423	185,268	241,921
	Dependent variable: Child mortality				
Odds Ratio	1.42**	1.45**	0.99	1.24	1.45***
		(0.231)	(0.256)	(0.162)	(0.172)
Marginal Effect		0.004**	0.000	0.002	0.002***
		(0.002)	(0.001)	(0.001)	(0.001)
N	138,063	150,845	34,423	185,268	192,456
	Dependent variable: Post-neonatal mortality				
Odds Ratio	1.42***	1.42***	1.14	1.30***	1.12*
		(0.136)	(0.187)	(0.105)	(0.075)
Marginal Effect		0.007***	0.001	0.004***	0.001
		(0.002)	(0.002)	(0.001)	(0.001)
N	138,063	150,845	34,423	185,268	226,142
	Dependent variable: Neonatal mortality				
Odds Ratio	1.23***	1.18**	1.46***	1.25***	1.27***
		(0.076)	(0.158)	(0.068)	(0.057)
Marginal Effect		0.006**	0.010***	0.007***	0.006***
		(0.002)	(0.003)	(0.002)	(0.001)
N	138,063	150,845	34,423	185,268	238,640

Note: Column (1) shows the odds ratio from logit regression in [67], while Column (2), (3) & (4) shows the odds ratio from logit regression with specification exactly the same as in [67]. The differences in odds ratio presented in Column (1) & (2) are due to difference in number of observations because I control for exactly the same variables as in [67] (including type of cooking fuel, place of residence, wealth index, mother's age, mother's education, mother's working status, sex of child, breastfeeding status, separate kitchen, type of house, and year of survey). Hence, I consider that [67] unnecessarily dropped about 12,000 observations. I have very few observations in Column (3) while the number of observations in NFHS-4 is more than that in first three rounds combined. The reason why I see this significant drop in number of observations is because only a (state module) sub-sample of women were asked about their employment status, resulting a lot of missing observations for mother's working status variable in the NFHS-4 (2015-16). In fact, most (82.5%) of the observations of this particular variable are missing for only the most recent survey. Column (5) copies the results from probit regressions presented in Tables 2.5–2.8 and presents odds ratio from logit regressions with my specification. One of my controls, a variable indicating whether household cooks inside the house, in a separate building, or outdoors, is only available in NFHS-4, thus, I essentially use only last round of the survey in Column (5). The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 2.10: Cooking Fuel Choice and Infant Mortality from IV (2SLS) Regressions (IVs = District × Year FEs)

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting	Under-Five	Child	Post-Neonatal	Neonatal
Polluting fuel for cooking		0.038***	0.001	0.002	0.035***
		(0.004)	(0.001)	(0.002)	(0.003)
Constant	-0.031	0.007**	0.005***	0.005***	-0.003
	(0.019)	(0.003)	(0.001)	(0.002)	(0.003)
Demographic controls	Y	Y	Y	Y	Y
N	242,971	242,971	242,971	242,971	242,971
R^2	0.51	0.02	0.00	0.01	0.01
F-stat on IVs	36.11				

Note: The first column reports result from the first-stage regression of 2SLS regression where the dependent variable is a binary variable for polluting fuel. Since I estimate 1,558 coefficients on IVs, I was unable to report the coefficient estimate on the instrumental variables. The F-test on IVs—the interactions of district dummies with year-of-interview fixed effect (District-by-Year FE)—for polluting fuel suggests that the instrument is statistically significant. Column (2), (3), (4) & (5) report results from the second-stage regressions of 2SLS regression with different dependent variable and similar specification where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 2.11: First-Stage Results on the Effect of Forest Cover on Cooking Fuel Choice

	Dependent variable: Polluting fuel for cooking				
	(1)	(2)	(3)	(4)	(5)
Forest cover	0.037***	0.008*	0.001	0.036***	0.036***
	(0.006)	(0.004)	(0.004)	(0.005)	(0.005)
Constant	0.722***	0.085***	0.063***	-0.024***	-0.027***
	(0.001)	(0.005)	(0.005)	(0.007)	(0.009)
Demographic controls	N	Y	Y	Y	Y
Year FE	N	N	Y	Y	Y
State FE	N	N	N	Y	Y
State-by-Year FE	N	N	N	N	Y
N	235,355	225,232	225,232	225,232	225,232
R^2	0.00	0.46	0.46	0.48	0.48
F-stat on IV	39.86	3.56	0.03	46.36	45.77

Note: The dependent variable is fuel type (polluting) for cooking. Each column reports OLS coefficient estimates on the IV–forest cover as a percent of total geographical area of the region measured at district-level. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The year fixed effects include year-of-interview dummies (2015 and 2016). State fixed effects include dummies for the states. Since variable of my interest, forest cover, is measured at district-level, I am not able to include District-by-Year FEs. If I include District-by-Year FEs, the impact of forest cover on fuel choice will be washed off. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 2.12: Cooking Fuel Choice and Infant Mortality from IV (2SLS) Regressions (IV = Agricultural Land Ownership)

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting	Under-Five	Child	Post-Neonatal	Neonatal
Polluting fuel for cooking		0.040**	0.001	0.007	0.033**
		(0.019)	(0.006)	(0.010)	(0.015)
Constant	-0.120***	0.011	0.005***	0.003	0.002
	(0.025)	(0.007)	(0.002)	(0.004)	(0.005)
Owns agricultural land	0.052***				
	(0.002)				
Demographic controls	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y
District-by-Year FE	Y	Y	Y	Y	Y
N	242,971	242,971	242,971	242,971	242,971
R^2	0.52	0.03	0.01	0.01	0.02
F-stat on IVs	694.92				

Note: The first column reports result from the first-stage regression of 2SLS regression where the dependent variable is a binary variable for polluting fuel. The F-test on IV—an indicator variable for household's agricultural land ownership—verifies that the instrument generates a plausible variation in polluting fuel for cooking. Column (2), (3), (4) & (5) report results from the second-stage regressions of 2SLS regression with different dependent variable and similar specification where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table 2.13: Cooking Fuel Choice and Infant Mortality from IV (2SLS) Regressions (IVs = Forest Cover & Agricultural Land Ownership)

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting	Under-Five	Child	Post-Neonatal	Neonatal
Polluting fuel for cooking		0.038**	0.003	-0.001	0.037**
		(0.018)	(0.006)	(0.009)	(0.015)
Constant	-0.139***	0.007	0.006***	0.002	-0.000
	(0.027)	(0.006)	(0.002)	(0.003)	(0.005)
Forest cover	0.037***				
	(0.009)				
Owns agricultural land	0.054***				
	(0.002)				
Demographic controls	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y
District-by-Year FE	Y	Y	Y	Y	Y
N	225,232	225,232	225,232	225,232	225,232
R^2	0.50	0.03	0.00	0.01	0.01
F-stat on IVs	354.36				
Hansen J statistic		2.14	0.72	3.44	1.13
χ^2 p-value		0.14	0.40	0.06	0.29

Note: The first column reports result from the first-stage regression of 2SLS regression where the dependent variable is a binary variable for polluting fuel. The F-test on IVs—district-wise forest cover measured as a percent of total geographical area and an indicator variable for household's agricultural land ownership—verifies that the instruments generate a plausible variation in polluting fuel for cooking. Column (2), (3), (4) & (5) report results from the second-stage regressions of 2SLS regression with different dependent variable and similar specification where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The Hansen's J-statistics suggest that the excluded IVs are exogenous and the model is not overidentified. The unit of observation is child. Parentheses contain standard errors clustered by PSUs. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

CHAPTER 3

A MODEL OF FUEL CHOICE AS A HOUSEHOLD PUBLIC GOOD

3.1 Introduction

While there is a limited but growing literature of rigorous empirical studies that attempt to identify the factors driving household fuel transition, theoretical foundation of household energy choice is also limited. A comprehensive survey of literature in conceptual and theoretical frameworks for household fuel choice/transition in developing countries have been reviewed by [65]. They conclude that much more research is required to better understand the channels of fuel transition and cooking fuel decision in spite of significant attention and growing empirical research on household fuel use.

Household choices of cooking fuel was initially determined by household income according to the "energy ladder" theory, heavily relying on traditional income effects in consumer model [15, 22, 46, 58, 59]. However, energy ladder model has been criticized by the fact that households in developing countries consume multiple fuels at the same time instead of completely transitioning from one fuel type to another [63, 83, 84]. As a result, the "fuel stacking" theory has been born and proposes that households use a combination of clean and polluting fuels with certain fraction. Consecutively, more sophisticated utility maximization models of urban and rural or agricultural households have been emerged to explain fuel choice patterns. First, [30] and [40] proposed urban households models in which utility maximizing consumers subject to budget constraint choose commercial and non-commercial fuels simultaneously conditional on fuel prices. Second, another wave of studies introduced market failures in theoretical framework of fuel choice [20, 47, 61]. The non-separated farm household models link the fuel choice with agricultural production and address rural labor market failures.

What type of fuel the household will use for cooking, on one hand, depends on the extent of household size or number of household members. This relationship is observed in my empirical chapter (Figure 2.4 in Chapter 2), indicating that large households are more likely to fire solid fuels to cook their food. The results from the first-stage regression of IV (2SLS) regression suggest that use of polluting fuel for cooking is positively associated with number of household members, supporting a positive relationship between household size and tendency of polluting fuel use shown in Figure 2.4.[1] My empirical results also confirm that household fuel choice should be endogenously determined within the model. In addition, the existing theoretical frameworks of fuel use do not explain the channels from household fuel choice to likelihood of child mortality. So the main objective of this chapter is to fill this important gap in the theoretical literature of household fuel use by linking fuel choice, a leading factor of indoor air pollution, with infant mortality.

In this chapter, motivated by my empirical findings in previous chapter and results from existing empirical studies, I present a model of overlapping generations to better understand the mechanism of household fuel choice decision and its association with infant mortality. In my theoretical framework, households maximize discounted utility by choice of the (i) investment level in a household level public good (e.g., clean cooking fuel) that benefits the survival likelihood as well as future earning potentials of each offspring, and (ii) fertility decisions

[1]The coefficient estimate on number of household members variable from the first-stage OLS regression where the dependent variable is a binary variable for polluting fuel was 0.016 (positive) and statistically significant at 1 percent level.

conditional on the likelihood of survival. I solve the household maximization problem for each family dynasty over the infinite horizon. Such an exercise will also enable us to articulate the reasons for government intervention, if any, to influence private investment in the household public goods.

The rest of the chapter is organized as follows. Section 3.2 describes a theoretical framework within which to examine the determination of household fuel choice and the potential for variation in cooking fuel decision to pass through into infant mortality or likelihood of survival. Section 3.3 solves the model, discusses interpretation of the model findings, and provides potential policy implications by elaborating on the comparative statics. Section 3.4 conducts a welfare analysis. Finally, Section 3.5 concludes.

3.2 Setup of the Model

Consider a model of overlapping generations, in which each individual lives for two-periods (adulthood and old age). Specifically, there are L number of family lines at each time period and each family at each period t has two members, adult and old age. In addition, the collection of family lines at a given time period t is referred to as a generation, while the collection of families across generations within a family line is referred to as a dynasty. Note that time is discrete, $t = 0, 1, \ldots$. Upon reaching adulthood, an individual at time t with earnings y_t:

- transfers σ fraction of his/her income to the parent, $t_t = \sigma y_t$;
- spends vn_t amount of income to raise n_t number of children, where v is the

cost per child;

- spends $k_t \in [0, \bar{k}]$ amount of income to purchase a household public good, which (i) increases the likelihood of survival of each of the n_t number of children (π_{t+1}) from π_0 to $\pi_0(1 + k_t)$,[2] and (ii) increases the earning capacity of surviving children (y_{t+1}) from y_0 to $y_0(1 + k_t)$.

The individual derives utility in period t from consumption, c_t, according to the Stone-Geary or Klein-Rubin utility function, U_t. I assume that the individual discounts each future period with a constant discount factor, $\beta < 1$. The discounted utility of an adult over the two-time period given a discount factor β therefore is

$$U_t = \log(c_t - \bar{c}) + \beta \log(c_{t+1} - \bar{c}), \qquad \beta \in (0, 1) \tag{3.1}$$

where $\bar{c}$ is a level of subsistence consumption. Since $y_t = y_0(1+k_{t-1})$, and the adult spends c_t, vn_t, k_t and σy_t respectively on consumption, child rearing, household public goods and transfer, the budget constraint of adult is:

$$y_0(1 + k_{t-1}) = c_t + vn_t + k_t + \sigma y_0(1 + k_{t-1}). \tag{3.2}$$

Since $\pi_{t+1} = \pi_0(1 + k_t)$ is the likelihood of survival of each of the n children, and $y_{t+1} = y_0(1 + k_t)$, the budget constraint of the individual during old age is:

$$\pi_{t+1} n_t \sigma y_{t+1} = c_{t+1} \tag{3.3}$$

[2]Since π_0 is the likelihood of child survival, then its complement, $1 - \pi_0$, is the likelihood of child mortality.

where the amount of transfers (only source of revenue for old age) that an old age will receive from each of his or her surviving child is $t_{t+1} = \sigma y_{t+1}$.

3.3 Solving the Model

I solve the following utility maximization problem of each adult surviving into adulthood at time t, by maximizing above-mentioned utility function subject to two budget constraints, taking as given his earning potential determined by investment undertaken by his parents k_{t-1}.[3]

$$\max_{k_t, n_t} U_t = \log(c_t - \bar{c}) + \beta \log(c_{t+1} - \bar{c})$$

subject to

$$y_0(1 + k_{t-1}) = c_t + v n_t + k_t + \sigma y_0(1 + k_{t-1}), \tag{3.4}$$

$$\pi_{t+1} n_t \sigma y_{t+1} = c_{t+1},$$

$$0 \leq k_t \leq \bar{k},$$

$$c_t \geq 0,\ y_0 \geq 0,\ k_{t-1} \geq 0,\ n_t \geq 0,\ \pi_{t+1} \geq 0.$$

Substituting the two budget constraints expressed in Equations (3.2) and (3.3) into the utility function in Equation (3.1) by determining c_t and c_{t+1}, we find

$$U_t = \underbrace{\log\left[y_0(1-\sigma)(1+k_{t-1}) - v n_t - k_t - \bar{c}\right]}_{\text{Adulthood}} + \underbrace{\beta \log\left(\pi_{t+1} n_t \sigma y_{t+1} - \bar{c}\right)}_{\text{Old age}} \tag{3.5}$$

[3]Detailed derivations of the first order conditions for this maximization problem are provided in Appendix B.

Maximizing U_t by choosing fertility n_t, we can write the first order condition as below, taking k_t as given for now:

$$\frac{v}{y_0(1-\sigma)(1+k_{t-1}) - vn_t - k_t - \bar{c}} = \frac{\beta\pi_{t+1}\sigma y_{t+1}}{\pi_{t+1}n_t\sigma y_{t+1} - \bar{c}} \tag{3.6}$$

and thus utility maximizing fertility level, n_t^*, is defined as

$$n_t^* = \frac{\beta y_0(1-\sigma)(1+k_{t-1})}{(1+\beta)v} - \frac{\beta k_t}{(1+\beta)v} - \frac{\beta\bar{c}}{(1+\beta)v} + \frac{\bar{c}}{(1+\beta)\pi_{t+1}\sigma y_{t+1}} \tag{3.7}$$

If we substitute the utility maximizing fertility level, n_t^*, back into the utility function in Equation (3.5), we derive the maximal discounted utility as

$$\begin{aligned} U_t(k_t, k_{t-1}) = (1+\beta)\log&\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1 + \frac{v}{\pi_{t+1}\sigma y_{t+1}}\right)\right] \\ &+\beta\log\left(\pi_{t+1}\sigma y_{t+1}\right) \\ &+\left\{\log\left(\frac{1}{1+\beta}\right) + \beta\log\left[\frac{\beta}{(1+\beta)v}\right]\right\} \end{aligned}$$

which is linearly proportional to the weighted average of the log of:

$$y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1 + \frac{v}{\pi_{t+1}\sigma y_{t+1}}\right)$$

and the log of

$$\pi_{t+1}\sigma y_{t+1}$$

plus a constant. I call this maximum $U_t(k_t, k_{t-1})$. Upon choosing the optimal level of fertility (or after internalizing the fertility decision), a consumer's (indirect) utility function or the consumer's maximal attainable utility depends on investment in clean cooking fuel in a quadratic form, i.e.,

$$
\begin{aligned}
U_t(k_t, k_{t-1}) = (1+\beta)\log\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1 + \frac{v}{\pi_0 y_0 \sigma (1+k_t)^2}\right)\right] \\
+\beta \log\left[\pi_0 y_0 \sigma (1+k_t)^2\right] \\
+\left\{\log\left(\frac{1}{1+\beta}\right) + \beta \log\left[\frac{\beta}{(1+\beta)v}\right]\right\}
\end{aligned} \tag{3.8}
$$

The comparative statics of the maximum $U_t(k_t, k_{t-1})$ with respect to k_t suggest several implications including:

- Investment in clean cooking fuel (or increase in k_t) would directly decrease an individual's utility during the adulthood age by reducing his/her disposable income for consumption; however, it can have a positive impact on utility over this period of life when:

$$
k_t \leq \left(\frac{2v\bar{c}}{\pi_0 y_0 \sigma}\right)^{1/3} - 1.
$$

- Investment in clean cooking fuel (or increase in k_t) increases an individual's utility over the old age period at all times via improving the child's likelihood of survival and resulting increase in transfer income.

$$
\frac{\partial \beta \log\left[\pi_0 y_0 \sigma (1+k_t)^2\right]}{\partial k_t} = \frac{2\beta}{1+k_t} > 0 \qquad \text{since } \beta > 0,\ k_t \geq 0.
$$

Now if we maximize $U_t(k_t, k_{t-1})$ in Equation (3.8) by choice of k_t, we can write the first order condition as

$$
\frac{(1+\beta)\left[1 - \frac{2v\bar{c}}{\pi_0 y_0 (1+k_t)^3}\right]}{y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left[1 + \frac{v}{\pi_0 y_0 (1+k_t)^2}\right]} = \frac{2\beta}{1+k_t} \tag{3.9}
$$

and derive the relationship between k_t and k_{t-1} as follows

$$k_{t-1} = \underbrace{\left[\frac{(1+3\beta)}{2\beta(1-\sigma)y_0}\right]}_{A} k_t - \underbrace{\left[\frac{1}{\beta(1-\sigma)\pi_0}\right]\left(\frac{v}{y_0}\right)\left(\frac{\bar{c}}{y_0}\right)}_{B} \frac{1}{(1+k_t)^2}$$
$$+ \underbrace{\frac{1}{1-\sigma}\left(\frac{\bar{c}}{y_0}\right) + \left[\frac{1+\beta}{2\beta(1-\sigma)}\right]\frac{1}{y_0} - 1}_{C}$$

or we can simplify this equation of motion, which illustrates the dynamic relation between k_t and k_{t-1}, by writing it in the following general form

$$k_{t-1} = Ak_t - \frac{B}{(1+k_t)^2} + C \tag{3.10}$$

Figure 3.1 shows the dynamic relationship between k_t and k_{t-1} expressed in Equation (3.10) under four different possibilities, which would depend on parameters. There are cases where k_t intersects with 45-degree line for multiple times (first three cases) or only one time at the origin (last case). Depending on when the k_t function intersects with 45-degree line, either before (Figure 3.1(a)) or after (Figure 3.1(b)) k_t curve reaches maximum level of investment in the household public good, the equilibrium outcomes are different. In the first two cases, (a) and (b), respectively we have less than $\bar{k}$ in $t-1$ being able to generate $\bar{k}$ in t, and conversely for the next case. In the third case, k_t function crosses 45-degree line exactly at $\bar{k}$; however, the equilibrium outcome of this case is similar to that of the first case (Figure 3.1(c)). Finally, there could be a case where k_t never intersects with 45^o line, except at the origin (Figure 3.1(d)).

In particular, there are total of three steady states or equilibria where $k_t = k_{t-1}$ in the first three cases. Since $k_t \geq 0$, $k_t = f(k_{t-1})$ starts from the origin where it intersects with the 45-degree line at $\tilde{k}_1$ for the first time. Then k_t function intersects with the 45-degree line at $\tilde{k}_2$ for the second time. Household's investment in public good is bounded, $k_t \in [0, \bar{k}]$, therefore, k_t function crosses the 45-degree line at $\tilde{k}_3$ for the last time ($\tilde{k}_3 \leq \bar{k}$). A steady state at $\tilde{k}_2$ is not dynamically stable since slope of $k_t = f(k_{t-1})$ function at $\tilde{k}_2$ is greater than 1. In each case, except for the last case, there are two long-run (stable) equilibria, one above and one below the unstable equilibrium at $\tilde{k}_2$. In other words, if k_{t-1} is less than the unstable equilibrium, dynamics will converge to zero.

3.4 Welfare Analysis

Each adult makes decisions to maximize his/her welfare, rather than that of the family dynasty. I call this "Holdup Problem", and the steady state welfare of an adult, where $k_{t-1} = k_t = k^*$, can be found solving the following equality for k^*:

$$(1+3\beta)\left[1 - y_0(1-\sigma)\right] k^{*3} + \left[2\beta\bar{c} + 7\beta + 3 - 3(1+3\beta)y_0(1-\sigma)\right] k^{*2} +$$

$$+ \left[4\beta\bar{c} + 5\beta + 3 - 3(1+3\beta)y_0(1-\sigma)\right] k^* -$$

$$- \underbrace{(1+3\beta)y_0(1-\sigma) + 2\beta\bar{c} + \beta + 1 - \frac{2v\bar{c}}{\pi_0 y_0 \sigma}}_{\text{Constant}} = 0$$

The above expression was found by replacing k_{t-1} and k_t with k^* in the maximal discounted utility, $U_t(k_t, k_{t-1})$, expressed in Equation (3.8), and maximized by choice of k^*. Detailed derivations are provided in Appendix C. Once I find the steady state welfare of an adult, we maximize the steady state welfare of the household to define under what condition will such k^* exceed (/is less than) the steady state equilibrium $\tilde{k}$ above. This will be further explored in the future work.

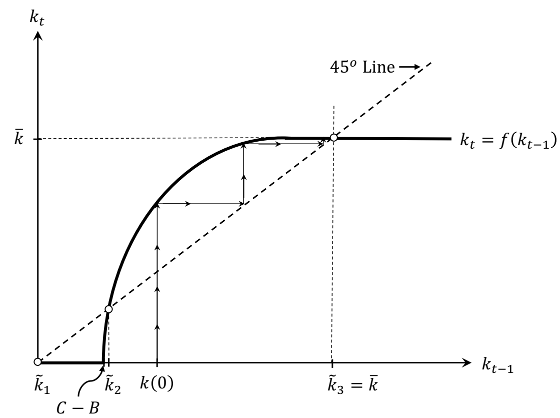
3.5 Conclusion

The overlapping generations (OLG) model is an ideal model setting in which the effect of fuel choice (which can be seen as household public good) on long-term earning potential and child mortality rate or inversely likelihood of child survival can be modeled in more structured way. Hence, I present a basic overlapping generations model with two-periods-lived individuals (adulthood and old age) who maximize the present discount value (PDV) of their own utility by controlling investment in a household level public good (clean cooking fuel, in this model) and fertility decision. The most significant contribution of my theoretical framework of fuel use to the currently growing literature is the fact that this is the first theory model which links the fuel use with infant mortality (or health outcome in general).

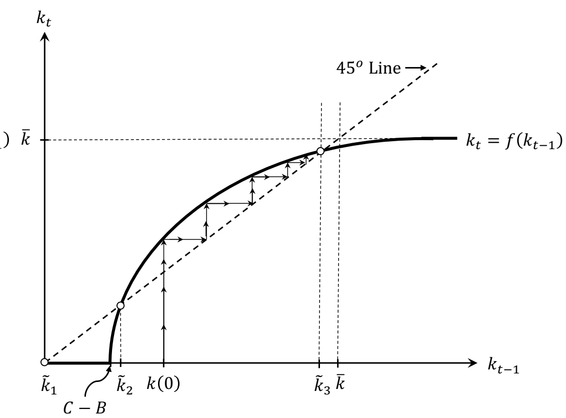
I utilized the Stone-Geary (Klein-Rubin) utility function to take into account subsistence consumption which is one of the most relevant features for developing countries. Fertility level is one of the two control or choice variables in my model, and I observe that the maximal discounted utility is linearly proportional to the weighted average of utilities gained over the adulthood and old age period after optimally choosing fertility level to maximize the utility. I also find that investment in clean fuel increases individual's utility over the old age period through increased survival likelihood of the child. On the other hand, individual's utility during adulthood period increases in response to investment in clean cooking fuel for some values of k_t (particularly, when k_t is less than certain value depending on parameters). There are two empirical implications from the model. First, the use of dirty fuel and fertility are positively correlated. Second, mortality rate and the use of dirty fuel are positively correlated.

Furthermore, upon optimally choosing investment in household public good, k_t, or maximizing an indirect utility function with respect to k_t, the dynamic relationship between k_{t-1} and k_t was found to be nonlinear. The convergence of OLG economy to steady state shows that there are multiple potential equilibrium outcomes at $\tilde{k} = \bar{k}$, $\tilde{k} < \bar{k}$, or $\tilde{k} = 0$. There is at least one steady-state equilibrium which is dynamically stable for $k_t = f(k_{t-1})$ function intersects the 45^o line. Finally, I formulate a problem, called as "Holdup Problem", to find the steady state welfare of an adult (k^*) and provide some solution strategy.

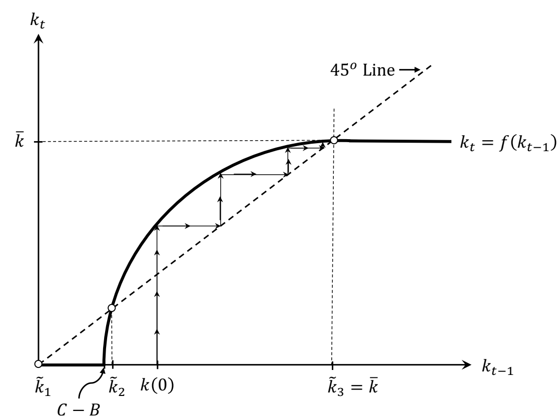
Figure 3.1: Convergence of OLG Economy to Steady State



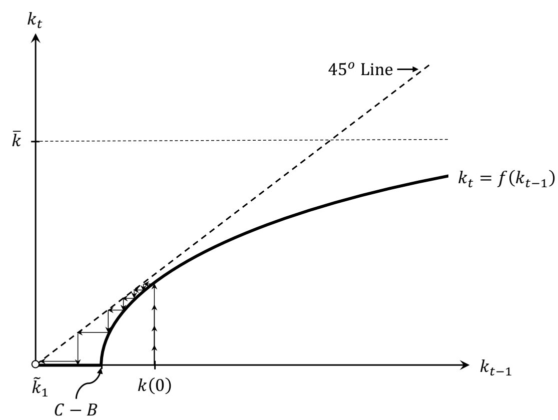
(a) k_t intersects 45^o line at the origin and after reaching $\bar{k}$



(b) k_t intersects 45^o line at the origin and before reaching $\bar{k}$



(c) k_t intersects 45^o line at the origin and exactly at $\bar{k}$



(d) k_t intersects 45^o line only at the origin

APPENDIX A

RESULTS FROM ROBUSTNESS CHECKS

Table A.1: Cooking Fuel Choice and Infant Mortality from IV Probit Regressions (IVs = District × Year FEs)

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting	Under-Five	Child	Post-Neonatal	Neonatal
Polluting fuel for cooking		0.412***	-0.087	-0.030	0.501***
		(0.057)	(0.134)	(0.094)	(0.068)
Constant	0.068***	-1.947***	-2.465***	-2.337***	-2.203***
	(0.006)	(0.036)	(0.082)	(0.058)	(0.043)
Demographic controls	Y	Y	Y	Y	Y
N	242,971	241,921	192,456	226,142	238,640
R^2	0.52				
F-stat on IVs	29.90				
Model Wald χ^2		4,832.40	619.06	1,277.82	2,959.27
Model degrees of freedom		17.00	17.00	17.00	17.00
Model Wald p-value		0.00	0.00	0.00	0.00
Exogeneity test Wald p-value		0.00	0.24	0.86	0.00
Wald χ^2 test of exogeneity		41.26	1.39	0.03	47.37

Note: The first column reports result from the first-stage OLS regression of IV probit where the dependent variable is a binary variable for polluting fuel. Since I estimate 1,558 coefficients on IVs, I was unable to report the coefficient estimates on the instrumental variables. The F-test on IVs—the interactions of district dummies with year-of-interview fixed effect (District-by-Year FE)—for polluting fuel suggests that the instrument is statistically significant. Column (2), (3), (4) & (5) report results from the IV probit regressions with different dependent variables and similar specification. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The unit of observation is child. Standard errors in parentheses are clustered at the PSU level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table A.2: Cooking Fuel Choice and Infant Mortality from IV Probit Regressions (IV = Agricultural Land Ownership)

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting	Under-Five	Child	Post-Neonatal	Neonatal
Polluting fuel for cooking		0.505**	0.151	0.256	0.596**
		(0.211)	(0.527)	(0.353)	(0.247)
Constant	-0.120***	-2.528***	-3.061***	-2.919***	-2.729***
	(0.025)	(0.215)	(0.339)	(0.325)	(0.255)
Owns agricultural land	0.052***				
	(0.002)				
Demographic controls	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y
District-by-Year FE	Y	Y	Y	Y	Y
N	242,971	241,921	192,456	226,142	238,640
R^2	0.52				
F-stat on IV	694.92				
Model Wald χ^2		6,445.99	981.37	2,057.59	4,231.22
Model degrees of freedom		618.00	433.00	547.00	604.00
Model Wald p-value		0.00	0.00	0.00	0.00
Exogeneity test Wald p-value		0.05	0.97	0.54	0.04
Wald χ^2 test of exogeneity		3.79	0.00	0.37	4.15

Note: The first column reports result from the first-stage OLS regression of IV probit where the dependent variable is a binary variable for polluting fuel. The F-test on IV—household's agricultural land ownership—confirms that the instrument creates a significant variation in polluting fuel for cooking. Column (2), (3), (4) & (5) report results from the IV probit regressions with different dependent variables and similar specification. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table A.3: Cooking Fuel Choice and Infant Mortality from IV Probit Regressions (IVs = Forest Cover & Agricultural Land Ownership)

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting	Under-Five	Child	Post-Neonatal	Neonatal
Polluting fuel for cooking		0.431**	0.216	-0.028	0.589***
		(0.198)	(0.452)	(0.329)	(0.228)
Constant	-0.139***	-2.599***	-2.811***	-2.944***	-2.812***
	(0.027)	(0.173)	(0.133)	(0.248)	(0.211)
Forest cover	0.037***				
	(0.009)				
Owns agricultural land	0.054***				
	(0.002)				
Demographic controls	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y
State-by-Year FE	Y	Y	Y	Y	Y
N	225,232	225,232	223,004	224,826	225,232
R^2	0.50				
F-stat on IVs	354.36				
Model Wald χ^2		5,467.66	772.22	1,460.91	3,519.84
Model degrees of freedom		52.00	46.00	51.00	52.00
Model Wald p-value		0.00	0.00	0.00	0.00
Exogeneity test Wald p-value		0.10	0.87	0.84	0.03
Wald χ^2 test of exogeneity		2.73	0.03	0.04	4.50

Note: The first column reports result from the first-stage OLS regression of IV probit where the dependent variable is a binary variable for polluting fuel. The F-test on IVs—district-wise forest cover measured as a percent of total geographical area of the region and an indicator variable for household's agricultural land ownership—confirms that the instruments create a significant variation in polluting fuel for cooking. Column (2), (3), (4) & (5) report results from the IV probit regressions with different dependent variables and similar specification. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table A.4: Cooking Fuel Choice and Infant Mortality from OLS Regressions

	(1) Under-Five	(2) Child	(3) Post-Neonatal	(4) Neonatal
Polluting fuel for cooking	0.008***	0.001***	0.001*	0.006***
	(0.001)	(0.000)	(0.001)	(0.001)
Constant	0.006	0.005***	0.003	-0.002
	(0.006)	(0.002)	(0.003)	(0.005)
Demographic controls	Y	Y	Y	Y
District-by-Year FE	Y	Y	Y	Y
N	242,971	242,971	242,971	242,971
R^2	0.03	0.01	0.01	0.02

Note: Each column reports the coefficient estimates from a multivariate OLS regression where the dependent variable is one of the four dummy variables for infant mortality and the key explanatory variable is indoor air pollution (polluting fuel for cooking). The demographic controls are exactly the same as those in Table 2.5–2.8. The interactions term include the interactions of dummies for 603 districts with year-of-interview fixed effects (District-by-Year FE). The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table A.5: Probit: The Marginal Impact of Cooking Fuel Choice on Infant Mortality (with Cookstoves Program States)

	(1) Under-Five	(2) Child	(3) Post-Neonatal	(4) Neonatal
Polluting fuel for cooking	0.008***	0.002***	0.001	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
Cookstoves Program States (NBCI)	0.057***	-0.005	0.008	0.043***
	(0.020)	(0.004)	(0.010)	(0.016)
Demographic controls	Y	Y	Y	Y
District-by-Year FE	Y	Y	Y	Y
N	241,921	192,456	226,142	238,640
Probit log-likelihood	-38,701	-6,225	-12,961	-27,562

Note: Each column reports average marginal effects for a multivariate probit regression where the dependent variable is one of the four dummy variables for infant mortality and the key explanatory variable is indoor air pollution (polluting fuel for cooking). In addition to demographic controls, I control for a dummy variable indicating states where National Biomass Cookstove Initiative (NBCI) has been implemented by the Government of India. The year fixed effects include dummies for years of interview. The district fixed effects include dummies for 603 districts, and the interactions term include the interactions of district dummies with year-of-interview fixed effects (District-by-Year FE). The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table A.6: Cooking Fuel Choice and Infant Mortality from IV (2SLS) Regressions (IVs = Forest Cover & Agricultural Land Ownership), with Cookstoves Program States

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting	Under-Five	Child	Post-Neonatal	Neonatal
Polluting fuel for cooking		0.038**	0.003	-0.001	0.037**
		(0.018)	(0.006)	(0.009)	(0.015)
Cookstoves Program States (NBCI)	0.056**	0.051***	0.006***	0.010***	0.035***
	(0.027)	(0.005)	(0.001)	(0.003)	(0.004)
Constant	-0.139***	0.007	0.006***	0.002	-0.000
	(0.027)	(0.006)	(0.002)	(0.003)	(0.005)
Forest cover	0.037***				
	(0.009)				
Owns agricultural land	0.054***				
	(0.002)				
Demographic controls	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y
District-by-Year FE	Y	Y	Y	Y	Y
N	225,232	225,232	225,232	225,232	225,232
R^2	0.50	0.03	0.00	0.01	0.01
F-stat on IVs	354.36				
Hansen J statistic		2.14	0.72	3.44	1.13
χ^2 p-value		0.14	0.40	0.06	0.29

Note: The first column reports result from the first-stage regression of 2SLS regression where the dependent variable is a binary variable for polluting fuel. The F-test on IVs—district-wise forest cover measured as a percent of total geographical area and an indicator variable for household's agricultural land ownership—verifies that the instruments generate a plausible variation in polluting fuel for cooking. Column (2), (3), (4) & (5) report results from the second-stage regressions of 2SLS regression with different dependent variable and similar specification where the key explanatory variable is the fitted value of polluting fuel from the first-stage estimation. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. In addition to these demographic controls, I control for a dummy variable indicating states where National Biomass Cookstove Initiative (NBCI) has been implemented by the Government of India. The Hansen's J-statistics suggest that the excluded IVs are exogenous, and the model is not overidentified. The unit of observation is child. Parentheses contain standard errors clustered by PSUs. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table A.7: Dirtiness of Cooking Fuels and Under-Five Mortality from OLS Regressions

	Dependent variable: Under-five mortality				
	(1)	(2)	(3)	(4)	(5)
Dirtiness level of cooking fuels	0.005***	0.002***	0.002***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.023***	0.030***	0.027***	0.002	0.007
	(0.001)	(0.004)	(0.004)	(0.006)	(0.006)
Demographic controls	N	Y	Y	Y	Y
Year FE	N	N	Y	Y	Y
State-by-Year FE	N	N	N	Y	N
District-by-Year FE	N	N	N	N	Y
N	354,161	230,091	230,091	230,091	230,091
R^2	0.00	0.02	0.02	0.03	0.03

Note: Each column reports the coefficient estimates from a multivariate OLS regression where the dependent variable is a dummy variable for under-five mortality. The key explanatory variable is a continuous variable that I created by ranking fuels used for cooking based on their dirtiness/cleanliness level. The cleanest fuel takes a value of 1 while the most polluting fuel takes a value of 10 (1 = electricity, 2 = LPG/natural gas, 3 = biogas, 4 = kerosene, 5 = coal/lignite, 6 = charcoal, 7 = wood, 8 = straw/shrubs/grass, 9 = agricultural crop, and 10 = animal dung). The demographic controls are exactly the same as those in Table 2.5–2.8. The year fixed effects include dummies for year-of-interview. The state and district fixed effects include dummies for 36 states and 603 districts, respectively, and the interactions term include the interactions of state and district dummies with year-of-interview fixed effects (State-by-Year FE and District-by-Year FE). The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table A.8: Dirtiness of Cooking Fuels and Child Mortality from OLS Regressions

	Dependent variable: Child mortality				
	(1)	(2)	(3)	(4)	(5)
Dirtiness level of cooking fuels	0.001***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.001***	0.008***	0.008***	0.005***	0.005***
	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)
Demographic controls	N	Y	Y	Y	Y
Year FE	N	N	Y	Y	Y
State-by-Year FE	N	N	N	Y	N
District-by-Year FE	N	N	N	N	Y
N	354,161	230,091	230,091	230,091	230,091
R^2	0.00	0.00	0.00	0.00	0.01

Note: Each column reports the coefficient estimates from a multivariate OLS regression where the dependent variable is a dummy variable for child mortality. The key explanatory variable is a continuous variable that I created by ranking fuels used for cooking based on their dirtiness/cleanliness level. The cleanest fuel takes a value of 1 while the most polluting fuel takes a value of 10 (1 = electricity, 2 = LPG/natural gas, 3 = biogas, 4 = kerosene, 5 = coal/lignite, 6 = charcoal, 7 = wood, 8 = straw/shrubs/grass, 9 = agricultural crop, and 10 = animal dung). The demographic controls are exactly the same as those in Table 2.5–2.8. The year fixed effects include dummies for year-of-interview. The state and district fixed effects include dummies for 36 states and 603 districts, respectively, and the interactions term include the interactions of state and district dummies with year-of-interview fixed effects (State-by-Year FE and District-by-Year FE). The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table A.9: Dirtiness of Cooking Fuels and Post-Neonatal Mortality from OLS Regressions

	Dependent variable: Post-neonatal mortality				
	(1)	(2)	(3)	(4)	(5)
Dirtiness level of cooking fuels	0.001***	0.000*	0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.007***	0.011***	0.010***	0.003	0.003
	(0.000)	(0.002)	(0.002)	(0.003)	(0.004)
Demographic controls	N	Y	Y	Y	Y
Year FE	N	N	Y	Y	Y
State-by-Year FE	N	N	N	Y	N
District-by-Year FE	N	N	N	N	Y
N	354,161	230,091	230,091	230,091	230,091
R^2	0.00	0.01	0.01	0.01	0.01

Note: Each column reports the coefficient estimates from a multivariate OLS regression where the dependent variable is a dummy variable for post-neonatal mortality. The key explanatory variable is a continuous variable that I created by ranking fuels used for cooking based on their dirtiness/cleanliness level. The cleanest fuel takes a value of 1 while the most polluting fuel takes a value of 10 (1 = electricity, 2 = LPG/natural gas, 3 = biogas, 4 = kerosene, 5 = coal/lignite, 6 = charcoal, 7 = wood, 8 = straw/shrubs/grass, 9 = agricultural crop, and 10 = animal dung). The demographic controls are exactly the same as those in Table 2.5–2.8. The year fixed effects include dummies for year-of-interview. The state and district fixed effects include dummies for 36 states and 603 districts, respectively, and the interactions term include the interactions of state and district dummies with year-of-interview fixed effects (State-by-Year FE and District-by-Year FE). The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table A.10: Dirtiness of Cooking Fuels and Neonatal Mortality from OLS Regressions

	Dependent variable: Neonatal mortality				
	(1)	(2)	(3)	(4)	(5)
Dirtiness level of cooking fuels	0.003***	0.002***	0.002***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.015***	0.010***	0.009***	-0.005	-0.001
	(0.001)	(0.003)	(0.003)	(0.004)	(0.005)
Demographic controls	N	Y	Y	Y	Y
Year FE	N	N	Y	Y	Y
State-by-Year FE	N	N	N	Y	N
District-by-Year FE	N	N	N	N	Y
N	354,161	230,091	230,091	230,091	230,091
R^2	0.00	0.01	0.01	0.02	0.02

Note: Each column reports the coefficient estimates from a multivariate OLS regression where the dependent variable is a dummy variable for neonatal mortality. The key explanatory variable is a continuous variable that I created by ranking fuels used for cooking based on their dirtiness/cleanliness level. The cleanest fuel takes a value of 1 while the most polluting fuel takes a value of 10 (1 = electricity, 2 = LPG/natural gas, 3 = biogas, 4 = kerosene, 5 = coal/lignite, 6 = charcoal, 7 = wood, 8 = straw/shrubs/grass, 9 = agricultural crop, and 10 = animal dung). The demographic controls are exactly the same as those in Table 2.5–2.8. The year fixed effects include dummies for year-of-interview. The state and district fixed effects include dummies for 36 states and 603 districts, respectively, and the interactions term include the interactions of state and district dummies with year-of-interview fixed effects (State-by-Year FE and District-by-Year FE). The unit of observation is child. Standard errors in parentheses are clustered at the primary sampling unit (PSU) level. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

Table A.11: Dirtiness of Cooking Fuels and Infant Mortality from IV (2SLS) Regressions (IVs = Forest Cover & Agricultural Land Ownership)

	1^{st} stage	2^{nd} stage			
	(1)	(2)	(3)	(4)	(5)
	Polluting	Under-Five	Child	Post-Neonatal	Neonatal
Dirtiness level of cooking fuels		0.006**	0.000	0.000	0.005**
		(0.003)	(0.001)	(0.001)	(0.002)
Constant	0.312***	0.002	0.005***	0.003	-0.006
	(0.118)	(0.006)	(0.002)	(0.003)	(0.005)
Forest cover	0.200***				
	(0.057)				
Owns agricultural land	0.357***				
	(0.013)				
Demographic controls	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
District FE	Y	Y	Y	Y	Y
District-by-Year FE	Y	Y	Y	Y	Y
N	212,493	212,493	212,493	212,493	212,493
R^2	0.52	0.03	0.00	0.01	0.02
F-stat on IVs	367.52				
Hansen J statistic		2.19	0.61	3.73	1.04
χ^2 p-value		0.14	0.44	0.05	0.31

Note: The first column reports result from the first-stage regression of 2SLS regression where the dependent variable is a continuous variable measuring the dirtiness of fuels used for cooking. The F-test on IVs—district-wise forest cover measured as a percent of total geographical area and an indicator variable for household's agricultural land ownership—verifies that the instruments generate a plausible variation in polluting fuel for cooking. Column (2), (3), (4) & (5) report results from the second-stage regressions of 2SLS regression with different dependent variable and similar specification where the key explanatory variable is the fitted value of fuel dirtiness from the first-stage estimation. The demographic controls include household characteristics: place of residence, household wealth, number of household members, place where food is cooked, and type of house; mother characteristics: age and education level; and child characteristics: gender and breastfeeding status. The Hansen's J-statistics suggest that the excluded IVs are exogenous and the model is not overidentified. The unit of observation is child. Parentheses contain standard errors clustered by PSUs. Significance: *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

APPENDIX B

A DETAILED SOLUTION OF THE MODEL

Recall that the utility maximization problem is defined as follows

$$\max_{k_t, n_t} U_t = \log(c_t - \bar{c}) + \beta \log(c_{t+1} - \bar{c})$$

subject to

$$y_0(1 + k_{t-1}) = c_t + v n_t + k_t + \sigma y_0(1 + k_{t-1}),$$

$$\pi_{t+1} n_t \sigma y_{t+1} = c_{t+1},$$

$$0 \leq k_t \leq \bar{k},$$

$$c_t \geq 0,\ y_0 \geq 0,\ k_{t-1} \geq 0,\ n_t \geq 0,\ \pi_{t+1} \geq 0.$$

To solve this maximization problem, I first find c_t and c_{t+1} from the two budget constraints as below

$$c_t = y_0(1 - \sigma)(1 + k_{t-1}) - v n_t - k_t,$$

$$c_{t+1} = \pi_{t+1} n_t \sigma y_{t+1}.$$

Plugging these two expressions into the utility function in Equation (3.1), we get

$$U_t = \underbrace{\log\left[y_0(1 - \sigma)(1 + k_{t-1}) - v n_t - k_t - \bar{c}\right]}_{\text{Adulthood}} + \underbrace{\beta \log\left(\pi_{t+1} n_t \sigma y_{t+1} - \bar{c}\right)}_{\text{Old age}}.$$

The first order condition with respect to fertility level n_t is

$$\frac{\partial U_t}{\partial n_t} = 0 \quad \Rightarrow \quad \frac{v}{y_0(1-\sigma)(1+k_{t-1}) - vn_t - k_t - \bar{c}} = \frac{\beta\pi_{t+1}\sigma y_{t+1}}{\pi_{t+1}n_t\sigma y_{t+1} - \bar{c}}$$

$$\begin{aligned} v\pi_{t+1}n_t\sigma y_{t+1} - v\bar{c} = &\ \beta\pi_{t+1}\sigma y_{t+1}y_0(1-\sigma)(1+k_{t-1}) - \beta\pi_{t+1}\sigma y_{t+1}vn_t \\ &-\beta\pi_{t+1}\sigma y_{t+1}k_t - \beta\pi_{t+1}\sigma y_{t+1}\bar{c} \end{aligned}$$

$$\begin{aligned} (1+\beta)v\pi_{t+1}n_t\sigma y_{t+1} = &\ \beta\pi_{t+1}\sigma y_{t+1}y_0(1-\sigma)(1+k_{t-1}) - \beta\pi_{t+1}\sigma y_{t+1}k_t \\ &-\beta\pi_{t+1}\sigma y_{t+1}\bar{c} + v\bar{c}. \end{aligned}$$

From this I can define the utility maximizing fertility level, n_t^*, as

$$\begin{aligned} n_t^* = &\ \frac{\beta\pi_{t+1}\sigma y_{t+1}y_0(1-\sigma)(1+k_{t-1})}{(1+\beta)v\pi_{t+1}\sigma y_{t+1}} - \frac{\beta\pi_{t+1}\sigma y_{t+1}k_t}{(1+\beta)v\pi_{t+1}\sigma y_{t+1}} \\ &-\frac{\beta\pi_{t+1}\sigma y_{t+1}\bar{c}}{(1+\beta)v\pi_{t+1}\sigma y_{t+1}} + \frac{v\bar{c}}{(1+\beta)v\pi_{t+1}\sigma y_{t+1}} \\ = &\ \frac{\beta y_0(1-\sigma)(1+k_{t-1})}{(1+\beta)v} - \frac{\beta k_t}{(1+\beta)v} - \frac{\beta\bar{c}}{(1+\beta)v} + \frac{\bar{c}}{(1+\beta)\pi_{t+1}\sigma y_{t+1}} \end{aligned}$$

To substitute the utility maximizing fertility level, n_t^*, back into the utility function in Equation (3.5), I will do this for adulthood and old age separately and then combine those two expressions together.

The adulthood part of the utility function is

$$\begin{aligned} c_t - \bar{c} = &\ y_0(1-\sigma)(1+k_{t-1}) - vn_t^* - k_t - \bar{c} \\ = &\ y_0(1-\sigma)(1+k_{t-1}) - \left(\frac{\beta}{1+\beta}\right)y_0(1-\sigma)(1+k_{t-1}) + \left(\frac{\beta}{1+\beta}\right)k_t + \left(\frac{\beta}{1+\beta}\right)\bar{c} \\ &-\frac{v\bar{c}}{(1+\beta)\pi_{t+1}\sigma y_{t+1}} - k_t - \bar{c} \end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{1}{1+\beta}\right) y_0(1-\sigma)(1+k_{t-1}) - \left(\frac{1}{1+\beta}\right) k_t - \left(\frac{1}{1+\beta}\right)\bar{c} - \frac{v\bar{c}}{(1+\beta)\pi_{t+1}\sigma y_{t+1}} \\
&= \left(\frac{1}{1+\beta}\right) y_0(1-\sigma)(1+k_{t-1}) - \left(\frac{1}{1+\beta}\right) k_t - \left(\frac{1}{1+\beta}\right)\bar{c}\left(1+\frac{v}{\pi_{t+1}\sigma y_{t+1}}\right) \\
&= \left(\frac{1}{1+\beta}\right)\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1+\frac{v}{\pi_{t+1}\sigma y_{t+1}}\right)\right]
\end{aligned}
$$

<u>Adulthood</u>: $\log(c_t - \bar{c}) = \log\left\{\left(\frac{1}{1+\beta}\right)\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1+\frac{v}{\pi_{t+1}\sigma y_{t+1}}\right)\right]\right\}$

$$
= \log\left(\frac{1}{1+\beta}\right) + \log\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1+\frac{v}{\pi_{t+1}\sigma y_{t+1}}\right)\right].
$$

The old age part of the utility function is

$$
\begin{aligned}
c_{t+1} - \bar{c} &= \pi_{t+1} n_t^* \sigma y_{t+1} - \bar{c} \\
&= \frac{(\pi_{t+1}\sigma y_{t+1})\beta y_0(1-\sigma)(1+k_{t-1})}{(1+\beta)v} - \frac{(\pi_{t+1}\sigma y_{t+1})\beta k_t}{(1+\beta)v} \\
&\quad - \frac{(\pi_{t+1}\sigma y_{t+1})\beta\bar{c}}{(1+\beta)v} + \frac{(\pi_{t+1}\sigma y_{t+1})\bar{c}}{(1+\beta)\pi_{t+1}\sigma y_{t+1}} - \bar{c} \\
&= \frac{(\pi_{t+1}\sigma y_{t+1})\beta y_0(1-\sigma)(1+k_{t-1})}{(1+\beta)v} - \frac{(\pi_{t+1}\sigma y_{t+1})\beta k_t}{(1+\beta)v} \\
&\quad - \frac{\beta}{(1+\beta)v}\left(\pi_{t+1}\sigma y_{t+1} + v\right)\bar{c} \\
&= \frac{\beta}{(1+\beta)v}\left[(\pi_{t+1}\sigma y_{t+1}) y_0(1-\sigma)(1+k_{t-1}) - (\pi_{t+1}\sigma y_{t+1})k_t - (\pi_{t+1}\sigma y_{t+1} + v)\,\bar{c}\right] \\
&= \left[\frac{\beta}{(1+\beta)v}\right](\pi_{t+1}\sigma y_{t+1})\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c} - \frac{v\bar{c}}{\pi_{t+1}\sigma y_{t+1}}\right] \\
&= \left[\frac{\beta}{(1+\beta)v}\right](\pi_{t+1}\sigma y_{t+1})\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1+\frac{v}{\pi_{t+1}\sigma y_{t+1}}\right)\right]
\end{aligned}
$$

<u>Old age</u>: $\beta\log(c_{t+1} - \bar{c}) = \beta\log\left[\frac{\beta}{(1+\beta)v}\right] + \beta\log(\pi_{t+1}\sigma y_{t+1})$

$$
+\beta\log\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1+\frac{v}{\pi_{t+1}\sigma y_{t+1}}\right)\right].
$$

Hence, using expressions for adulthood and old age, we can write the maximal discounted utility as below

$$
\begin{aligned}
U_t(k_t, k_{t-1}) &= \log(c_t - \bar{c}) + \beta \log(c_{t+1} - \bar{c}) \\
&= \log\left(\frac{1}{1+\beta}\right) + \log\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1 + \frac{v}{\pi_{t+1}\sigma y_{t+1}}\right)\right] \\
&\quad + \beta \log\left[\frac{\beta}{(1+\beta)v}\right] + \beta \log(\pi_{t+1}\sigma y_{t+1}) \\
&\quad + \beta \log\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1 + \frac{v}{\pi_{t+1}\sigma y_{t+1}}\right)\right] \\
&= \underbrace{(1+\beta)\log\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1 + \frac{v}{\pi_{t+1}\sigma y_{t+1}}\right)\right]}_{D} \\
&\quad + \underbrace{\beta \log\left(\pi_{t+1}\sigma y_{t+1}\right)}_{E} \\
&\quad + \underbrace{\left\{\log\left(\frac{1}{1+\beta}\right) + \beta \log\left[\frac{\beta}{(1+\beta)v}\right]\right\}}_{\text{Constant } (F)}.
\end{aligned}
$$

Now I maximize this maximum $U_t(k_t, k_{t-1})$ by choice of k_t, and let's derive the first order condition (FOC) with respect to k_t below

$$
\begin{aligned}
D &= (1+\beta)\log\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left(1 + \frac{v}{\pi_{t+1}\sigma y_{t+1}}\right)\right] \\
&= (1+\beta)\log\left[y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c} - \frac{v\bar{c}}{\pi_{t+1}\sigma y_{t+1}}\right] \\
\frac{\partial D}{\partial k_t} &= \frac{(1+\beta)\left[\frac{2v\bar{c}}{\pi_0 y_0 (1+k_t)^3} - 1\right]}{y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left[1 + \frac{v}{\pi_0 y_0 (1+k_t)^2}\right]}
\end{aligned}
$$

and

$$E = \beta \log(\pi_{t+1}\sigma y_{t+1}) = \beta \log\left[\pi_0 y_0 (1+k_t)^2\right]$$

$$\frac{\partial E}{\partial k_t} = \frac{2\beta\pi_0 y_0(1+k_t)}{\pi_0 y_0 (1+k_t)^2} = \frac{2\beta}{1+k_t}$$

therefore, the FOC is

$$\frac{\partial U_t(k_t, k_{t-1})}{\partial k_t} = \frac{\partial D}{\partial k_t} + \frac{\partial E}{\partial k_t} + \underbrace{\frac{\partial F}{\partial k_t}}_{=0} = 0$$

$$\frac{(1+\beta)\left[\frac{2v\bar{c}}{\pi_0 y_0(1+k_t)^3} - 1\right]}{y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left[1 + \frac{v}{\pi_0 y_0(1+k_t)^2}\right]} + \frac{2\beta}{1+k_t} = 0$$

$$\frac{(1+\beta)\left[1 - \frac{2v\bar{c}}{\pi_0 y_0(1+k_t)^3}\right]}{y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left[1 + \frac{v}{\pi_0 y_0(1+k_t)^2}\right]} = \frac{2\beta}{1+k_t}.$$

By rearranging and collecting the terms as below, I define the equation of motion which demonstrates the dynamic relationship between k_t and k_{t-1}.

$$(1+\beta)(1+k_t) - \frac{2(1+\beta)v\bar{c}}{\pi_0 y_0(1+k_t)^2} = 2\beta\left\{y_0(1-\sigma)(1+k_{t-1}) - k_t - \bar{c}\left[1 + \frac{v}{\pi_0 y_0(1+k_t)^2}\right]\right\}$$

$$(1+\beta)(1+k_t) - \left[\frac{2(1+\beta)v\bar{c}}{\pi_0 y_0}\right]\frac{1}{(1+k_t)^2} = 2\beta y_0(1-\sigma)(1+k_{t-1}) - 2\beta k_t - 2\beta\bar{c} - \left(\frac{2\beta v\bar{c}}{\pi_0 y_0}\right)\frac{1}{(1+k_t)^2}$$

$$(1+\beta)(1+k_t) - (1+\beta)\left(\frac{2v\bar{c}}{\pi_0 y_0}\right)\frac{1}{(1+k_t)^2} + \beta\left(\frac{2v\bar{c}}{\pi_0 y_0}\right)\frac{1}{(1+k_t)^2} + 2\beta k_t = 2\beta y_0(1-\sigma)(1+k_{t-1}) - 2\beta\bar{c}$$

$$1 + k_t + \beta + \beta k_t + 2\beta k_t - \left(\frac{2v\bar{c}}{\pi_0 y_0}\right)\frac{1}{(1+k_t)^2} = 2\beta y_0(1-\sigma)(1+k_{t-1}) - 2\beta\bar{c}$$

$$(1+3\beta)k_t - \left(\frac{2v\bar{c}}{\pi_0 y_0}\right)\frac{1}{(1+k_t)^2} = 2\beta y_0(1-\sigma)k_{t-1} + 2\beta y_0(1-\sigma) - 2\beta\bar{c} - \beta - 1$$

$$2\beta y_0(1-\sigma)k_{t-1} = (1+3\beta)k_t - \left(\frac{2v\bar{c}}{\pi_0 y_0}\right)\frac{1}{(1+k_t)^2} - 2\beta y_0(1-\sigma) + 2\beta\bar{c} + \beta + 1$$

and if we divide the both sides of above expression by $2\beta y_0(1-\sigma)$, we get the following equation of motion between k_t and k_{t-1}

$$\begin{aligned} k_{t-1} = &\underbrace{\left[\frac{(1+3\beta)}{2\beta(1-\sigma)y_0}\right]}_{A} k_t - \underbrace{\left[\frac{1}{\beta(1-\sigma)\pi_0}\right]\left(\frac{v}{y_0}\right)\left(\frac{\bar{c}}{y_0}\right)}_{B}\frac{1}{(1+k_t)^2} \\ &+ \underbrace{\frac{1}{1-\sigma}\left(\frac{\bar{c}}{y_0}\right) + \left[\frac{1+\beta}{2\beta(1-\sigma)}\right]\frac{1}{y_0} - 1}_{C} \end{aligned}$$

or

$$k_{t-1} = Ak_t - \frac{B}{(1+k_t)^2} + C.$$

APPENDIX C

A DETAILED DERIVATION OF THE WELFARE ANALYSIS

By replacing k_{t-1} and k_t with k^* in the maximal discounted utility, $U_t(k_t, k_{t-1})$, expressed in Equation (3.8), we find

$$U(k^*) = (1+\beta)\log\left[y_0(1-\sigma)(1+k^*) - k^* - \bar{c}\left(1 + \frac{v}{\pi_0 y_0 \sigma (1+k^*)^2}\right)\right]$$
$$+\beta \log\left[\pi_0 y_0 \sigma (1+k^*)^2\right]$$
$$+\left\{\log\left(\frac{1}{1+\beta}\right) + \beta\log\left[\frac{\beta}{(1+\beta)v}\right]\right\}$$

We maximize the utility function with respect to k^* and found the following first order condition:

$$\frac{\partial U(k^*)}{\partial k^*} = 0 \quad \Rightarrow \quad \frac{(1+\beta)\left[y_0(1-\sigma) - 1 + \left(\frac{2v\bar{c}}{\pi_0 y_0 \sigma}\right)\frac{1}{(1+k^*)^3}\right]}{y_0(1-\sigma)(1+k^*) - k^* - \bar{c} - \frac{v\bar{c}}{\pi_0 y_0 \sigma (1+k^*)^2}} + \frac{2\beta\pi_0 y_0 \sigma (1+k^*)}{\pi_0 y_0 \sigma (1+k^*)^2} = 0$$

$$\frac{(1+\beta)\left[1 - y_0(1-\sigma) - \left(\frac{2v\bar{c}}{\pi_0 y_0 \sigma}\right)\frac{1}{(1+k^*)^3}\right]}{y_0(1-\sigma)(1+k^*) - k^* - \bar{c} - \frac{v\bar{c}}{\pi_0 y_0 \sigma (1+k^*)^2}} = \frac{2\beta}{1+k^*}$$

$$(1+\beta)\left[1 - y_0(1-\sigma)\right](1+k^*) - (1+\beta)\left(\frac{2v\bar{c}}{\pi_0 y_0 \sigma}\right)\frac{1}{(1+k^*)^2} = 2\beta y_0(1-\sigma)(1+k^*) -$$
$$-2\beta k^* - 2\beta\bar{c} - \frac{2\beta v\bar{c}}{\pi_0 y_0 \sigma (1+k^*)^2}$$

$$\left[1 - y_0(1-\sigma) + \beta - 3\beta y_0(1-\sigma)\right](1+k^*) = \left(\frac{2v\bar{c}}{\pi_0 y_0 \sigma}\right)\frac{1}{(1+k^*)^2} - 2\beta k^* - 2\beta\bar{c}$$

$$\left[1 - y_0(1-\sigma) + \beta - 3\beta y_0(1-\sigma)\right] + \left[1 - y_0(1-\sigma) + \beta - 3\beta y_0(1-\sigma)\right]k^* =$$
$$= \left(\frac{2v\bar{c}}{\pi_0 y_0 \sigma}\right)\frac{1}{(1+k^*)^2} - 2\beta k^* - 2\beta\bar{c}$$

$$\left[1 - y_0(1-\sigma) + \beta - 3\beta y_0(1-\sigma) + 2\beta\right] k^* = \left(\frac{2v\bar{c}}{\pi_0 y_0 \sigma}\right) \frac{1}{(1+k^*)^2} +$$
$$+ \underbrace{y_0(1-\sigma) + 3\beta y_0(1-\sigma)}_{(1+3\beta)y_0(1-\sigma)} - 2\beta\bar{c} - \beta - 1$$

$$\left[1 - y_0(1-\sigma) + 3\beta - 3\beta y_0(1-\sigma)\right] k^* = \left(\frac{2v\bar{c}}{\pi_0 y_0 \sigma}\right) \frac{1}{(1+k^*)^2} + (1+3\beta)y_0(1-\sigma) - 2\beta\bar{c} - \beta - 1$$

$$(1+3\beta)\left[1 - y_0(1-\sigma)\right] k^* = \left(\frac{2v\bar{c}}{\pi_0 y_0 \sigma}\right) \frac{1}{(1+k^*)^2} + (1+3\beta)y_0(1-\sigma) - 2\beta\bar{c} - \beta - 1$$

$$(1+3\beta)\left[1 - y_0(1-\sigma)\right] k^*(1+k^*)^2 = \frac{2v\bar{c}}{\pi_0 y_0 \sigma} + \left[(1+3\beta)y_0(1-\sigma) - 2\beta\bar{c} - \beta - 1\right](1+k^*)^2$$

$$(1+3\beta)\left[1 - y_0(1-\sigma)\right] k^* \left(1 + 2k^* + k^{*2}\right) = \left[(1+3\beta)y_0(1-\sigma) - 2\beta\bar{c} - \beta - 1\right] \cdot$$
$$\cdot \left(1 + 2k^* + k^{*2}\right) + \frac{2v\bar{c}}{\pi_0 y_0 \sigma}$$

By rearranging the above expression and collecting the terms, I get the following equation:

$$(1+3\beta)\left[1 - y_0(1-\sigma)\right] k^{*3} + \left[2\beta\bar{c} + 7\beta + 3 - 3(1+3\beta)y_0(1-\sigma)\right] k^{*2} +$$
$$+ \left[4\beta\bar{c} + 5\beta + 3 - 3(1+3\beta)y_0(1-\sigma)\right] k^* -$$
$$- \underbrace{(1+3\beta)y_0(1-\sigma) + 2\beta\bar{c} + \beta + 1 - \frac{2v\bar{c}}{\pi_0 y_0 \sigma}}_{\text{Constant}} = 0.$$

BIBLIOGRAPHY

[1] II Abubakar, T Tillmann, and A Banerjee. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the global burden of disease study 2013. *Lancet*, 385(9963):117–171, 2015.

[2] Julia C Allen and Douglas F Barnes. The causes of deforestation in developing countries. *Annals of the association of American Geographers*, 75(2):163–184, 1985.

[3] Komalkirti Apte and Sundeep Salvi. Household air pollution and its effects on health. *F1000Research*, 5, 2016.

[4] Shams Arifeen, Robert E Black, Gretchen Antelman, Abdullah Baqui, Laura Caulfield, and Stan Becker. Exclusive breastfeeding reduces acute respiratory infection and diarrhea deaths among infants in Dhaka slums. *Pediatrics*, 108(4):e67–e67, 2001.

[5] Deepak Bajracharya. Fuel, food or forest? Dilemmas in a Nepali village. *World Development*, 11(12):1057–1074, 1983.

[6] Kalpana Balakrishnan, Sagnik Dey, Tarun Gupta, RS Dhaliwal, Michael Brauer, Aaron J Cohen, Jeffrey D Stanaway, Gufran Beig, Tushar K Joshi, Ashutosh N Aggarwal, et al. The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: The global burden of disease study 2017. *The Lancet Planetary Health*, 2018.

[7] Manuel Barron and Maximo Torero. Household electrification and indoor air pollution. *Journal of Environmental Economics and Management*, 86:81–92, 2017.

[8] Diego G Bassani, Prabhat Jha, Neeraj Dhingra, and Rajesh Kumar. Child mortality from solid-fuel use in India: A nationally-representative case-control study. *BMC Public Health*, 10(1):491, 2010.

[9] Prashant Bharadwaj, Matthew Gibson, Joshua Graff Zivin, and Christopher Neilson. Gray matters: Fetal pollution exposure and human capital formation. *Journal of the Association of Environmental and Resource Economists*, 4(2):505–542, 2017.

[10] Robert E Black, Saul S Morris, and Jennifer Bryce. Where and why are 10 million children dying every year? *The Lancet*, 361(9376):2226–2234, 2003.

[11] Branko Boskovic, Ujjayant Chakravorty, Martino Pelli, and Anna Risch. The effect of forest access on the market for fuelwood in India. Technical Report, CESifo Working Paper, 2018.

[12] Nigel Bruce, John McCracken, Rachel Albalak, Morten Scheid, Kirk R Smith, Victorina Lopez, and Chris West. Impact of improved stoves, house construction and child location on levels of indoor air pollution exposure in young Guatemalan children. *Journal of Exposure Science and Environmental Epidemiology*, 14(S1):S26, 2004.

[13] Nigel Bruce and Yasmin Von Schirnding. *Addressing the links between indoor air pollution, household energy and human health*. World Health Organization, 2002.

[14] Zoë A Chafe, Michael Brauer, Zbigniew Klimont, Rita Van Dingenen, Sumi Mehta, Shilpa Rao, Keywan Riahi, Frank Dentener, and Kirk R Smith. Household cooking with solid fuels contributes to ambient $PM_{2.5}$ air pollution and the burden of disease. *Environmental Health Perspectives*, 122(12):1314, 2014.

[15] Muyeye Chambwera and Henk Folmer. Fuel switching in harare: An almost ideal demand system approach. *Energy Policy*, 35(4):2538–2548, 2007.

[16] Tom Chang, Joshua Graff Zivin, Tal Gross, and Matthew Neidell. Particulate pollution and the productivity of pear packers. *American Economic Journal: Economic Policy*, 8(3):141–69, 2016.

[17] Tom Y Chang, Joshua Graff Zivin, Tal Gross, and Matthew Neidell. The effect of pollution on worker productivity: Evidence from call center workers in China. *American Economic Journal: Applied Economics*, 11(1):151–72, 2019.

[18] Kenneth Y Chay and Michael Greenstone. Air quality, infant mortality, and the clean air act of 1970. Technical Report, National Bureau of Economic Research, 2003.

[19] Kenneth Y Chay and Michael Greenstone. The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *The Quarterly Journal of Economics*, 118(3):1121–1167, 2003.

[20] Le Chen, Nico Heerink, and Marrit van den Berg. Energy consumption in rural china: A household model for three villages in Jiangxi province. *Ecological Economics*, 58(2):407–420, 2006.

[21] Alice H Cushing, Jonathan M Samet, William E Lambert, Betty J Skipper, William C Hunt, Stephen A Young, and Leroy C McLaren. Breastfeeding reduces risk of respiratory illness in infants. *American Journal of Epidemiology*, 147(9):863–870, 1998.

[22] Sukanya Das, Hugo De Groote, Bhagirath Behera, et al. Determinants of household energy use in Bhutan. *Energy*, 69:661–672, 2014.

[23] Susmita Dasgupta, Mainul Huq, M Khaliquzzaman, Kiran Pandey, and David R Wheeler. Indoor air quality for poor families: New evidence from Bangladesh. *Indoor Air*, 16(6):426–444, 2006.

[24] Patrícia Muniz de Medeiros, Taline Cristina da Silva, Alyson Luiz Santos de Almeida, and Ulysses Paulino de Albuquerque. Socio-economic predictors of domestic wood use in an Atlantic forest area (North-East Brazil): A tool for directing conservation efforts. *International Journal of Sustainable Development & World Ecology*, 19(2):189–195, 2012.

[25] Esperanza Diaz, Tone Smith-Sivertsen, Dan Pope, Rolv T Lie, Anaite Diaz, John McCracken, Byron Arana, Kirk R Smith, and Nigel Bruce. Eye discomfort, headache and back pain among Mayan Guatemalan women taking part in a randomised stove intervention trial. *Journal of Epidemiology & Community Health*, 61(1):74–79, 2007.

[26] Douglas W Dockery, C Arden Pope, Xiping Xu, John D Spengler, James H Ware, Martha E Fay, Benjamin G Ferris Jr, and Frank E Speizer. An association between air pollution and mortality in six U.S. cities. *New England Journal of Medicine*, 329(24):1753–1759, 1993.

[27] Esther Duflo, Michael Greenstone, and Rema Hanna. Cooking stoves, indoor air pollution and respiratory health in rural Orissa. *Economic and Political Weekly*, pages 71–76, 2008.

[28] Esther Duflo, Michael Greenstone, and Rema Hanna. Indoor air pollution, health and economic well-being. *SAPI EN. S. Surveys and Perspectives Integrating Environment and Society*, (1.1), 2008.

[29] Esther Duflo and Rema Hanna. Monitoring works: Getting teachers to

come to school. Technical Report, National Bureau of Economic Research, 2005.

[30] John HY Edwards and Christian Langpap. Startup costs and the decision to switch from firewood to gas fuel. *Land Economics*, 81(4):570–586, 2005.

[31] MB Epstein, MN Bates, NK Arora, K Balakrishnan, DW Jack, and KR Smith. Household fuels, low birth weight, and neonatal death in India: The separate impacts of biomass, kerosene, and coal. *International Journal of Hygiene and Environmental Health*, 216(5):523–532, 2013.

[32] Osita Kingsley Ezeh, Kingsley Emwinyore Agho, Michael John Dibley, John Joseph Hall, and Andrew Nicolas Page. The effect of solid fuel use on childhood mortality in Nigeria: Evidence from the 2013 cross-sectional household survey. *Environmental Health*, 13(1):113, 2014.

[33] Deon Filmer and Lant Pritchett. Estimating wealth effects without expenditure data–or tears: An application to educational enrollments in states of India. *Demography*, 38(1):115–132, 2001.

[34] Elizabeth Frankenberg, Douglas McKee, and Duncan Thomas. Health consequences of forest fires in Indonesia. *Demography*, 42(1):109–129, 2005.

[35] Joshua Graff Zivin and Matthew Neidell. The impact of pollution on worker productivity. *American Economic Review*, 102(7):3652–73, 2012.

[36] Joshua Graff Zivin and Matthew Neidell. Environment, health, and human capital. *Journal of Economic Literature*, 51(3):689–730, 2013.

[37] Joshua Graff Zivin and Matthew Neidell. Air pollution's hidden impacts. *Science*, 359(6371):39–40, 2018.

[38] William Greene. Discrete choice modeling. In *Palgrave Handbook of Econometrics*, pages 473–556. Springer, 2009.

[39] Michael Greenstone and Rema Hanna. Environmental regulations, air and water pollution, and infant mortality in India. *American Economic Review*, 104(10):3038–72, 2014.

[40] Gautam Gupta and Gunnar Köhlin. Preferences for domestic fuel: analysis with socio-economic factors and rankings in Kolkata, India. *Ecological Economics*, 57(1):107–121, 2006.

[41] Emily S Gurley, Nusrat Homaira, Henrik Salje, Pavani K Ram, Rashidul Haque, William Petri, Joseph Bresee, William J Moss, Patrick Breysse, Stephen P Luby, et al. Indoor exposure to particulate matter and the incidence of acute lower respiratory infections among children: A birth cohort study in urban Bangladesh. *Indoor Air*, 23(5):379–386, 2013.

[42] RD Hanbar and Priyadarshini Karve. National programme on improved chulha (NPIC) of the government of India: An overview. *Energy for Sustainable Development*, 6(2):49–55, 2002.

[43] Rema Hanna, Esther Duflo, and Michael Greenstone. Up in smoke: the influence of household behavior on the long-run impact of improved cooking stoves. *American Economic Journal: Economic Policy*, 8(1):80–114, 2016.

[44] Rema Hanna and Paulina Oliva. The effect of pollution on labor supply: Evidence from a natural experiment in Mexico city. *Journal of Public Economics*, 122:68–79, 2015.

[45] M Jane Heinig. Host defense benefits of breastfeeding for the infant: Effect of breastfeeding duration and exclusivity. *Pediatric Clinics of North America*, 48(1):105–123, 2001.

[46] Rasmus Heltberg. Factors determining household fuel choice in Guatemala. *Environment and Development Economics*, 10(3):337–361, 2005.

[47] Rasmus Heltberg, Thomas Channing Arndt, and Nagothu Udaya Sekhar. Fuelwood consumption and forest degradation: A household model for domestic energy substitution in rural India. *Land Economics*, pages 213–232, 2000.

[48] John P Holdren, Kirk R Smith, Tord Kjellstrom, David Streets, Xiaodong Wang, and S Fischer. Energy, the environment and health. *New York: United Nations Development Programme*, 2000.

[49] Imelda. Indoor air pollution and infant mortality: A new approach. In *AEA Papers and Proceedings*, volume 108, pages 416–421. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2018.

[50] Adam Isen, Maya Rossin-Slater, and W Reed Walker. Every breath you take–every dollar you'll make: The long-term consequences of the clean air act of 1970. *Journal of Political Economy*, 125(3):848–902, 2017.

[51] Maulik Jagnani. Poor sleep: Sunset time and human capital production, 2018.

[52] Seema Jayachandran. Air quality and early-life mortality evidence from Indonesia's wildfires. *Journal of Human Resources*, 44(4):916–954, 2009.

[53] Saori Kashima, Takashi Yorifuji, Toshihide Tsuda, Juliani Ibrahim, and Hiroyuki Doi. Effects of traffic-related outdoor air pollution on respiratory illness and mortality in children, taking into account indoor air pollution, in Indonesia. *Journal of Occupational and Environmental Medicine*, 52(3):340–345, 2010.

[54] M Khalequzzaman, M Kamijima, K Sakai, NA Chowdhury, N Hamajima, and T Nakajima. Indoor air pollution and its impact on children under five years old in Bangladesh. *Indoor Air*, 17(4):297–304, 2007.

[55] Md Khalequzzaman, Michihiro Kamijima, Kiyoshi Sakai, Takeshi Ebara, Bilqis Amin Hoque, and Tamie Nakajima. Indoor air pollution and health of children in biomass fuel-using households of Bangladesh: Comparison between urban and rural areas. *Environmental Health and Preventive Medicine*, 16(6):375–383, 2011.

[56] Md Khalequzzaman, Michihiro Kamijima, Kiyoshi Sakai, Bilqis Amin Hoque, and Tamie Nakajima. Indoor air pollution and the health of children in biomass-and fossil-fuel users of Bangladesh: Situation in two different seasons. *Environmental Health and Preventive Medicine*, 15(4):236–243, 2010.

[57] VVN Kishore and PV Ramana. Improved cookstoves in rural India: how improved are they?: A critique of the perceived benefits from the national programme on improved chulhas (NPIC). *Energy*, 27(1):47–63, 2002.

[58] Reza Kowsari and Hisham Zerriffi. Three dimensional energy profile: A conceptual framework for assessing household energy use. *Energy Policy*, 39(12):7505–7517, 2011.

[59] Jann Lay, Janosch Ondraczek, and Jana Stoever. Renewables in the energy transition: Evidence on solar home systems and lighting fuel choice in Kenya. *Energy Economics*, 40:350–359, 2013.

[60] Gwenaelle Legros, Ines Havet, Nigel Bruce, Sophie Bonjour, Kamal Rijal, Minoru Takada, et al. The energy access situation in developing coun-

tries: A review focusing on the least developed countries and sub-Saharan Africa. *World Health Organization and UNDP*, 2009.

[61] D. T. Manning and J. E. Taylor. Migration and fuel use in rural Mexico. *Ecological Economics*, 102(1):126–136, 2014.

[62] William J Martin, John W Hollingsworth, and Veerabhadran Ramanathan. Household air pollution from cookstoves: Impacts on health and climate. In *Global Climate Change and Public Health*, pages 237–255. Springer, 2014.

[63] Omar R Masera, Barbara D Saatkamp, and Daniel M Kammen. From linear fuel switching to multiple cooking strategies: A critique and alternative to the energy ladder model. *World Development*, 28(12):2083–2103, 2000.

[64] Premlata Menon. Indoor spatial monitoring of combustion generated pollutants (TSP, CO, and BaP) by Indian cookstoves. Technical Report, Hawaii Univ Honolulu Dept of Meteorology, 1988.

[65] Christophe Muller and Huijie Yan. Household fuel use in developing countries: Review of theory and evidence. *Energy Economics*, 70:429–439, 2018.

[66] Sabrina Naz, Andrew Page, and Kingsley Emwinyore Agho. Household air pollution and under-five mortality in Bangladesh (2004–2011). *International Journal of Environmental Research and Public Health*, 12(10):12847–12862, 2015.

[67] Sabrina Naz, Andrew Page, and Kingsley Emwinyore Agho. Household air pollution and under-five mortality in India (1992–2006). *Environmental Health*, 15(1):54, 2016.

[68] Shanta Pandey and Yuan Lin. Adjusted effects of domestic violence, tobacco use, and indoor air pollution from use of solid fuel on child mortality. *Maternal and Child Health Journal*, 17(8):1499–1507, 2013.

[69] Carolina Perez-Heydrich, Joshua L Warren, Clara R Burgert, and Michael Emch. *Guidelines on the use of DHS GPS data*. ICF International, 2013.

[70] AG Pinto, P Khanna, and CN Pandey. Gujarat wood balance study 1984. *Gujarat State Forest Department, Vadodara*, 1985.

[71] CA Pope, David V Bates, and Mark E Raizenne. Health effects of particu-

late air pollution: Time for reassessment? *Environmental Health Perspectives*, 103(5):472, 1995.

[72] EA Rehfuess, NG Bruce, and KR Smith. Solid fuel use: Health effect. In: Nriagu JO (ed.) Encyclopedia of environmental health, v 5, pp. 150161 Burlington: Elsevier, 2011. *Environmental Health*, 5:150–161, 2011.

[73] Eva A Rehfuess, Lia Tzala, Nicky Best, David J Briggs, and Mike Joffe. Solid fuel use and cooking practices as a major risk factor for ALRI mortality among African children. *Journal of Epidemiology & Community Health*, pages 887–892, 2009.

[74] Seppo T Rinne, Edgar J Rodas, Mikael L Rinne, Joshua M Simpson, and Larry T Glickman. Use of biomass fuel is associated with infant mortality and child health in trend analysis. *The American Journal of Tropical Medicine and Hygiene*, 76(3):585–591, 2007.

[75] S. Saksena, R. Prasad, RC Pal, and V. Joshi. Patterns of daily exposure to TSP and CO in the Garhwal Himalaya. *Atmospheric Environment. Part A. General Topics*, 26(11):2125–2134, 1992.

[76] Mirjam Schindler, Geoffrey Caruso, and Pierre Picard. Equilibrium and first-best city with endogenous exposure to local air pollution from traffic. *Regional Science and Urban Economics*, 62:12–23, 2017.

[77] Kirk R. Smith. National burden of disease in India from indoor air pollution. *Proceedings of the National Academy of Sciences*, 97(24):13286–13293, 2000.

[78] Tone Smith-Sivertsen, Esperanza Diaz, Dan Pope, Rolv T Lie, Anaite Diaz, John McCracken, Per Bakke, Byron Arana, Kirk R Smith, and Nigel Bruce. Effect of reducing indoor air pollution on women's respiratory symptoms and lung function: The respire randomized trial, Guatemala. *American Journal of Epidemiology*, 170(2):211–220, 2009.

[79] Maria Joana Specht, Severino Rodrigo Ribeiro Pinto, Ulysses Paulino Albuquerque, Marcelo Tabarelli, and Felipe PL Melo. Burning biodiversity: Fuelwood harvesting causes forest degradation in human-dominated tropical landscapes. *Global Ecology and Conservation*, 3:200–209, 2015.

[80] James M Tielsch, Joanne Katz, Ravilla D Thulasiraj, Christian L Coles, S Sheeladevi, Elizabeth L Yanik, and Lakshmi Rahmathullah. Exposure to

indoor biomass fuel and tobacco smoke and risk of adverse reproductive outcomes, mortality, respiratory morbidity and growth among newborn infants in south India. *International Journal of Epidemiology*, 38(5):1351–1363, 2009.

[81] UNICEF. *The state of the world's children 2007: Women and Children – The Double Dividend of Gender Equality*, volume 7. Unicef, 2006.

[82] Ashish Kumar Upadhyay, Abhishek Singh, Kaushalendra Kumar, and Ashish Singh. Impact of indoor air pollution from the use of solid fuels on the incidence of life threatening respiratory illnesses in children in India. *BMC Public Health*, 15(1):300, 2015.

[83] Bianca van der Kroon, Roy Brouwer, and Pieter Jh Van Beukering. The energy ladder: Theoretical myth or empirical truth? Results from a meta-analysis. *Renewable and Sustainable Energy Reviews*, 20:504–513, 2013.

[84] Bianca van der Kroon, Roy Brouwer, and Pieter JH van Beukering. The impact of the household decision environment on fuel choice behavior. *Energy Economics*, 44:236–247, 2014.

[85] WHO. Indoor air pollution: National burden of disease estimates. 2007.

[86] WHO. Indoor air quality guidelines: Household fuel combustion. 2014.

[87] WHO. Burning opportunity: Clean household energy for health, sustainable development, and wellbeing of women and children. 2016.

[88] WHO. *Inheriting a sustainable world? Atlas on children's health and the environment*. World Health Organization, 2017.

[89] WHO. Burden of disease from household air pollution for 2016. 2018.

[90] WHO. Household air pollution and health. *Fact sheet*, 292, 2018.

[91] WHO. World health statistics 2018: Monitoring health for the SDGs, sustainable development goals. 2018.

[92] J Wichmann and KVV Voyi. Influence of cooking and heating fuel use on 1–59 month old mortality in South Africa. *Maternal and Child Health Journal*, 10(6):553–561, 2006.

[93] Junfeng Jim Zhang and Kirk R Smith. Household air pollution from coal and biomass fuels in China: Measurements, health impacts, and interventions. *Environmental Health Perspectives*, 115(6):848, 2007.